Project Report – Dillon Constantine

# GitHub URL

<https://github.com/dconstantine22/UCDPA_DillonConstantine>

# Abstract

This report is an exploratory data analysis into petrol prices worldwide. With the petrol prices crisis that is currently unfolding right here in Ireland there was an opportunity to further understand the problem at hand worldwide. The report looks at trends in the data, looking into the top oil consumers, most affordable petrol prices worldwide and the countries that have the most gross domestic product per capita. The study took shape using an open-source petrol prices dataset that was readily available for use. The analysing was carried out using various python libraries such as Pandas and NumPy and visualisation libraries; matplotlib, seaborn and plotly.

# Introduction

The current inflation crisis is affecting people worldwide the prices of electricity, gas and in particular petrol and diesel are higher than ever before. This study investigates the trends of petrol prices worldwide from June 2022 and highlights the main insights from the exploratory data analysis.

The main problem statements that will be investigated throughout this study are;

* Visualise the highest & lowest oil consuming countries.
* Investigate the most expensive prices of petrol worldwide.
* Visualise the countries with the highest GDP Per Capita.
* Find the country with the biggest Yearly Gallon Consumption Per Capita
* Using linear regression predict the ‘GPD Per Capita’ based on the ‘Price Per Gallon’.

# Dataset

The dataset that was used in this project contained information on the petrol/oil prices worldwide (Usmani, 2022). The open-source dataset was collected by Zeeshan Ul Hassan Usmani and is available on Kaggle.com, an online collaborative sharing platform (Uslu, 2022). According to the metadata, (Usmani, 2022), of this dataset it was collected using Google API from public sources such as International Monetary Fund (IMF), the United Nations and World Bank. Note throughout this report oil and petrol are the key words referenced when talking about the data.

The original data set included 11 fields in total and 181 rows of data.

The fields were;

* Serial Number (Used as the Index Column)
* Country
* Daily Oil Consumption in Barrels
* World Share of Oil Consumption
* Yearly Gallon Per Capita
* Price Per Gallon in USD
* Price Per Liter in USD
* Price Per Liter in PKR (Pakistani Rupee)
* Gross Domestic Product (GDP) Per Capita
* Gallons GDP Per Capita Can Buy
* Number Times the GDP Per Capita can be Purchased of Petrol

The reasons behind the selection of this dataset were simple. The price of petrol/oil is a current topic, so it is on the mind of a lot of people. With the prices continuously rising right here in Ireland (O'Sullivan, 2022), looking into price trends worldwide was of interest. The dataset contained a good sample size with 181 countries included, clearly defined fields and the data was up to date as of June 2022.

# Implementation Process

This section of the report with step through the process in which the exploratory data analysis of Worldwide Petrol Prices took place from beginning to end.

## Step 1: Importing the Libraries

After selecting the dataset, the python packages that were going to be used throughout this study needed to be installed and imported. There were five libraries used in this study NumPy and Pandas to work with the data and then visualisation libraries such as Matplotlib, Seaborn and Plotly.

The following is some background on the selection of the libraries used and their main purpose in this exploratory analysis of petrol prices worldwide.

* The NumPy library is used within Python to work with numerical data (Developers, 2008-2022).
* The Pandas library is used for data analysis to generate data structures that make it easy and intuitive when working with different data types (Development, 2008-2022).
* The Matplotlib and Seaborn packages are used in this study to generate visual graphs to depict the data while the use of Plotly allows for the creation of interactive graphs.

## Step 2: Importing the Dataset

The dataset selected in this study was a csv file so using the Pandas library was ideal. Pandas has a function called read\_csv () which imported the data into a pandas data structure called df\_ppww (Data Frame Petrol Prices Worldwide).

Initially a ‘*UnicodeDecodeError’* error occurred when importing the dataset without any parameters within the pd.read\_csv("Petrol Dataset June 23 2022 -- Version 2.csv").

This error was due to the data requiring a larger width encoding parameter to be used such as ‘latin-1’ or ‘ISO-8859-1’. The ‘latin-1’ encoding parameter was used in this study. In the import statement the ‘Serial Number’ (S#) field was also defined as the index column as it contained unique values for each row of data (i.e., Each country).

Table

Description automatically generated

## Step 3: Data Understanding & Cleaning

It was important to look behind the data being used in this study. As it was a real-world data set there was potential to be flaws in the data.

### Shape and Data Types

To start the shape and data types of each field were established. A conversion was required on some object fields to convert them to ints and floats. The reason of the data type being object was due to special characters (‘%’, ‘,’) being included in the data. Included in this stage was a quick fix on spelling mistakes in the field names. Once the conversion and column name fixes were carried out a check was done to confirm the change.

### Create / Drop Fields

The next stage of this process was to create any new fields for analysis or drop unnecessary fields. To investigate the ‘daily liter price’ per country a new column was needed. This column was generated by taking the Daily Oil Consumption which was in barrels and multiplying by the quantity of litters in a barrel, 159L according to (Termeer, 2013, p. 4), and then multiply by the price per liter.

*(Daily Oil Consumption (Barrels) \* 159 Liters) \* Price Per Liter (USD)*

Also, in this stage an unnecessary field to this study was dropped from the dataset, Price Per Liter (PKR). The reason behind this was this study uses the United States Dollar (USD) throughout making the price per liter in Pakistan Rupees an unnecessary field.

### Null & Unique Value Check

A null value check was carried out next to check for any missing data. This check came back clear, and all data was accounted for as seen in the below figure.

A picture containing table

Description automatically generated

Another check was carried out to check for duplicate entries in the data and again this came back clear there was no duplications of the represented countries. Evidence of this can be seen in the figure below.

Table

Description automatically generated

### Descriptive Statistics

The final part of this step was to return the basic descriptive statistics of the petrol prices data. The output of this gives a clearer picture of the data as the min, max and mean of each field is returned as well as other useful information.

For instance, some of the basic stats of the Petrol Prices Worldwide date are;

* The Average Price for a Liter of Petrol: $1.50
* The Max Price for a Liter of Petrol: $14.50
* The Min Oil Consumed Daily: 51 Barrels
* The Average Oil Consumed Daily: 500,000 Barrels

## Step 4: Visualisation

Once the Data was cleaned, prepped and basic stats understood it was time to visualise the data. In this study there was three python libraries used as explained prior in this report, Matplotlib, Seaborn and Plotly. Through visualisations the trends within the data could be easily displayed using bar charts, histograms, and scatter plots. The trends and results of the EDA will be discussed in the results section of this report below.

# Results

Throughout this study of the data various trends were depicted on graphs and the results can be seen in this section of the report.

### Price Per Litre – Data Dispersion

The figure below shows the count of the Price Per Liter at each price point from the Petrol Prices Worldwide dataset. The graph shows us that most countries have a liter petrol price lower then $3. There are two countries above the rest with a liter of petrol price point being greater the $4.

Chart, histogram

Description automatically generated

## Correlation Graph

### Heat Map of all Field Correlations

The Heatmap below gives a good depiction of the various relationships between fields in the Petrol Prices Worldwide data. Corelations of 0.95 and above can be suspicious and with more data collection these relationships could be investigated further. Corelations of between 0.5 – 0.9 show a strong relationship between fields and this can be seen in the heatmap below.

A picture containing text, monitor, screen, several

Description automatically generated

Based off the above heat map there is a strong corelation of 0.622 between Yearly Gallons Per Capita and GDP Per Capita (USD). In the below figure it can be seen that many of the data points are along a similar area of the Scatter plot showing a strong relationship between the two fields.

Chart, scatter chart

Description automatically generated

## Petrol/Oil Prices

### Most expensive countries to buy Petrol/Oil

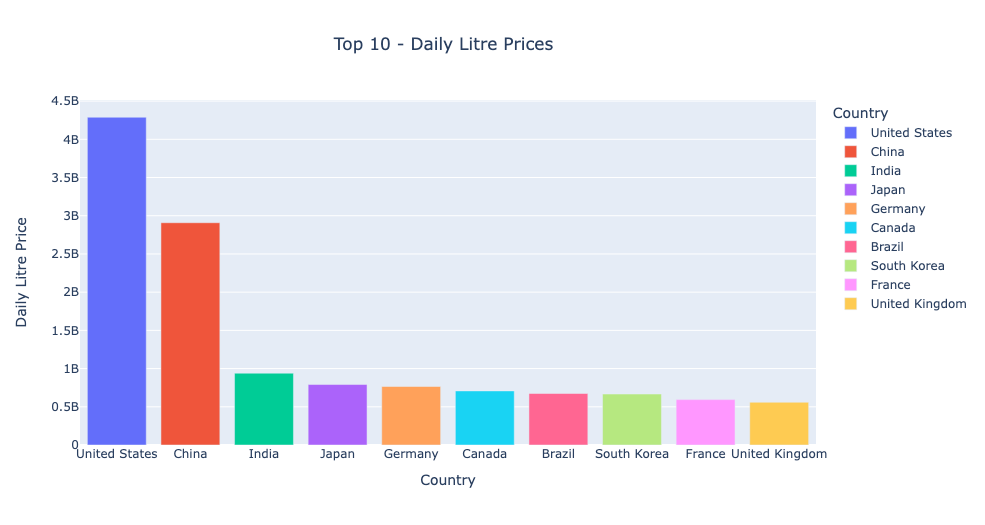
The graph below shows us the top 10 counters where the price of a litre of petrol is the greatest. From the graph we can see that North Korea’s petrol price is $14.50 this is nearly three times greater than the next highest price of Tonga. The graph also shows us how similar the petrol price is in many countries for example Hong Kong, Norway & Iceland all in $3 a litre price range.

Chart

Description automatically generated with medium confidence

### Daily Prices Petrol/Oil

The graph below shows the daily litre price per country. This graph considers the oil consumption and the price of a litre to create the daily litre price in the country. The United States and China are much greater than the rest with India behind. This may be due to the oil consumption which will be investigated in the next section of this report.



## Petrol/Oil Consumption

### Top Oil Consumers

The graph below represents the top 15 oil consumers worldwide. The bar chart shows us that United States and China are the top 2 consumers with the India and Japan also high consumers. The likely hood is this is due to the population and manufacturing outputs of these countries. According to (Global Upside, 2022), the US and China have the top 2 manufacturing outputs in the world which would burn plenty of oil. This also relates to the daily liter price as we seen above US and China were the top 2.

A picture containing bar chart

Description automatically generated

### Lowest Oil Consumers

On the other end of the scale the graph below represents the lowest consumers of oil in the world. Niue and Saint Helena consume little to no oil compared to that of US and China. This would come down to population size. From previous results it is also noted that Niue and Tonga are among the top 3 in the world for the litre price of petrol which result in lower usage.

Chart, bar chart

Description automatically generated

## Yearly Gallon Consumed Per Capita

The below graph represents the countries with the highest individual consumption. The picture this graph paints it that the individual people of Singapore consume a significantly more amount of petrol each year then other countries worldwide.

Chart, bubble chart

Description automatically generated

## Countries with the Highest GDP Per Capita

The graph below shows us that Luxemburg have the highest economic output of any nation worldwide per person. While US have the highest oil usage their economic output is 7th worldwide and China are not even in the top 10.

Chart, bubble chart

Description automatically generated

## Linear Regression Model

The linear regression model below uses Price Per Litre as the independent variable and the GDP Per Capita as the dependant variable. The line of best fit on the graph measures how the GDP per capita varies depending on the petrol price. The result is not accurate and requires more data to improve accuracy.

Chart

Description automatically generated with medium confidence

# Insights

Throughout this EDA of the petrol price worldwide data the main insights have been;

* Most petrol prices worldwide are lower then $3 a litre. With the main proportion price being approx. $1.80.
* The most expensive places to buy petrol in the world are North Korea ($14.50/litre), Tonga ($4.28/litre), Niue ($3.02/litre) and Hong Kong ($3/litre).
* The top 2 oil consumers in the world are the US and China with both consuming over 10 million barrels daily. This consumption comes from the population sizes of both nations and the fact they are the top manufacturing producers worldwide. This consumption compared to the lowest consumers of oil, Niue and Saint Helena, with 51 and 70 barrels respectively. These countries are on the low end of the scale due to their small populations.
* Leading the way with the highest yearly individual consumption of Oil is Singapore with 3679.5 Gallons and closest countries behind are less then 2000 gallons per year.
* The Highest Gross Domestic Product per capita is Luxemburg followed closely by 4 other countries Switzerland, Macao, Ireland, and Cayman Islands. These top 5 countries for economic output show they are prospering nations with petrol prices currently.
* The linear regression model used in this study needs more data to become more accurate. So, no accurate modelling could be done on this dataset.

# Bibliography

Developers, N., 2008-2022. *NumPy: the absolute basics for beginners.* [Online]   
Available at: https://numpy.org/doc/stable/user/absolute\_beginners.html  
[Accessed 16 July 2022].

Development, P., 2008-2022. *Pandas Package overview.* [Online]   
Available at: https://pandas.pydata.org/docs/getting\_started/overview.html#package-overview  
[Accessed July 2022].

Global Upside, L., 2022. *Top 10 Manufacturing Countries in the World.* [Online]   
Available at: https://globalupside.com/top-10-manufacturing-countries-in-the-world/  
[Accessed 19 July 2022].

O'Sullivan, J., 2022. *Cost of diesel and petrol up more than 40% on last year.* [Online]   
Available at: https://www.rte.ie/news/2022/0614/1304877-fuel-prices/  
[Accessed 17 July 2022].

Termeer, C., 2013. *Fundamentals of Investing in Oil and Gas.* illustrated ed. s.l.:Chris Termeer.

Uslu, C., 2022. *What is Kaggle?.* [Online]   
Available at: https://www.datacamp.com/blog/what-is-kaggle  
[Accessed July 2022].

Usmani, Z., 2022. *Petrol/Gas Prices Worldwide.* [Online]   
Available at: https://www.kaggle.com/datasets/zusmani/petrolgas-prices-worldwide  
[Accessed 11 July 2022].