Project Report

# GitHub URL

https://github.com/DiarmuidBarrett/UCDPA\_diarmuidbarrett.git

# Abstract

I have imported data from various sources on the price of oil in different regions (Europe and Global data), with the price being shown in two different formats (price per barrel and price per cubic meter).   
I have also imported data on daily shipping rates for three different classes of ship.

I then use various methods to reconcile my oil price data, ensuring it is consistent and thus reliable, while also performing analysis on the shipping prices to see how they are correlated to the price of oil, and each other.

I then use summary statistics and plot graphs to derive valuable insights from my analysis.

# Introduction

Working as a financial analyst, I find a lot of my work involves reconciling data to ensure it is consistent with what is expected, and consistent across all systems and records. Having also worked in a shipping company for a number of years, albeit some time back, this is data I knew I could use, was interested to work on, and a way to see if I could put my new Python skills to the test to use this raw data to derive valuable insights.

I wanted to see if I could use Pyhton as a way to reconcile various databases to ensure the information is consistent across both, and wanted to use different measure i.e. cubic meter and barrel prices, as in my current role I need to reconcile figures across different databases, in different currencies.

In the same way I knew what rate to convert barrel to cubic meters at, in my job as a financial analyst I will know what the past spot fx rates are and so in a sense I used the different measure of oil price as a proxy for different currencies, and different sources of information as a proxy for our different databases.

# Dataset

I have a number of datasets. I have European brent oil barrel prices in USD, from Reuters. I have Global crude oil prices in USD, per cubic meter, from “Our World in Data” (Our World in Data, 2022). I have shipping price data from Clarkson Research Services.

As mentioned above, I used this data as it would be a proxy for what I will be using Python for in my current role, and testing my skills on data I knew I could derive valuable insights from, should I have the skills to do so.

# Implementation Process

Firstly, I loaded in the relevant brent oil price data, per barrel in USD, from an xls file. Next, I stored it as a pandas data frame, removed the irrelevant rows using a .iloc formula, set the first row as the title and then re-indexed the data frame. I then set the index to be the date, set the oil price to be a number and used a groupby function to group the daily price data into years, using the mean price for the year. I then imported the crude oil price data, per cubic meter in USD, from a CSV file. Again, I removed the irrelevant rows at the top, and some irrelevant columns using the .iloc function, and the irrelevant rows at the bottom using the .tail function, finally I reindexed the data frame. Next, I merged the two oil price data frames into one yearly summary data frame. I then renamed the columns to be more simply to remember and type out and re-ordered them so that the year would be column 1.

The next thing I did was define a custom function so that I could quickly calculate brent oil price per barrel as a percentage of crude oil price per cubic meter so I could randomly spot check years to see if they % was close to target. One oil barrel is equal to roughly 0.159 cubic meters, so I wanted to quickly check my mean price per barrel for each year was close to 15.9% the price of the cubic meter price for each year. This served as a way for me to reconcile the two datasets to ensure they were at least closely aligned to one another and the information I had imported into my notebook was correct and usable.

Next, I decided I would add this percentage as a new column so I could print the data frame and see it for all years if I so wanted. To do this I converted it to a numpy array and used the np.column\_stack() function to add a column showing this percentage for each year and converted back into a pandas data frame.

I then imported my daily ship price information form an xlsx file, ensuring not to port a title in this time using the header = None function. I then cleaned the data in a similar way to with the oil prices, using .iloc and renamed the columns. I then filled in the blank values as ‘0’ using the fillna() function.

I then set the columns up to have Year as the first column, as can be seen from my script this took some trial and error and began my data analysis. I could not get the graphs to generate any values, so I ran a check, with some help form Google, and it showed me where I had an error. To fix this, I cleaned the tables some more, so I was working with similar time ranges and created multiple graphs using Matplotlib. I also ran some summary statistics on these for assistance in deriving valuable insights.

**Milestone 7 – Machine Learning**Apologies, I did not know where else to put this.

1. Given the historical data I have on oil prices and shipping rates, I could use machine learning and deep learning to predict future values of time series such as oil prices or, in this case, shipping rates.
2. I would use regression methods, for example, seeing how oil prices have changed over time, how they are currently behaving and using regression analysis to predict how this will drive/affect future shipping rates.

# Results

I first created a couple of charts to see how oil price would be correlated to daily shipping rates.   
A graph of a line graph

Description automatically generated with medium confidence

This first chart was to display how the mean daily price of the 310K tonne ship was correlated with the mean price of oil for each year. I will discuss the effectiveness if this in the insights section.

Next, I expanded on this graph and added in the price data for the two other shipping methods, not only to see how they were also correlated with the price of oil, bit to get a comparison of how the price of the different ships fluctuated over time. This graph can be seen here:

A graph showing the price of oil

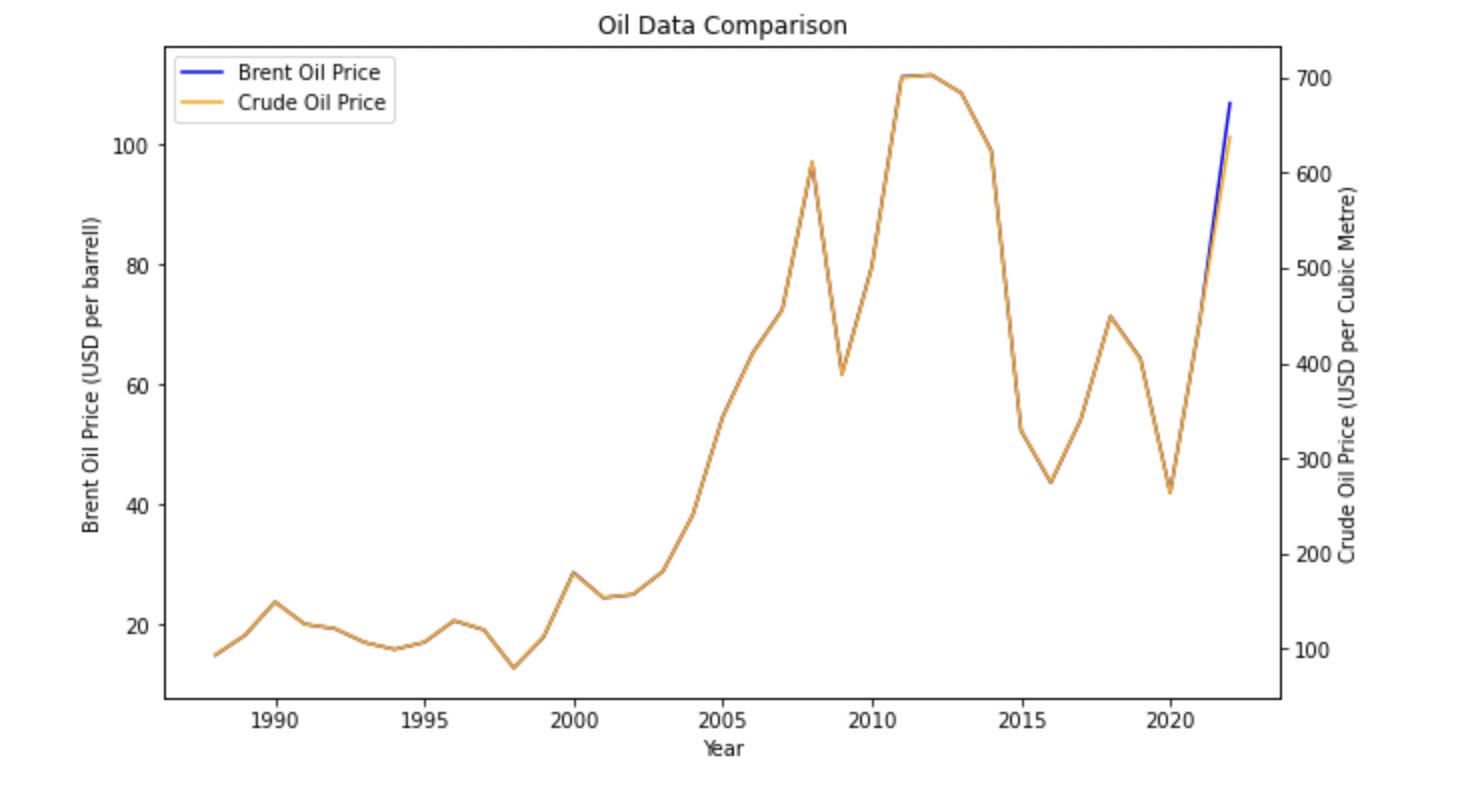
Description automatically generated

Finally, I also created graphs to show how the oil price data from my two sources aligned, and I believe this very accurately shows, despite having two different measure, the prices do reconcile and follow close on identical trends. I also calculated the correlation coefficient on this which I will mention further in my insights. See graph here:

A graph showing the price of oil

Description automatically generated

# Insights

* The first insight that I took is that the European data I extracted from the excel file (daily rates), are almost identical to the global data I extracted as a CVS file, meaning the European brent oil pricing data can be used as a good proxy for global oil prices. As mentioned above, the global data was extracted in cubic meters and the European data was received in barrels. A quick google search will tell us that the 1 barrel is equal to about 0.159 cubic meters. I used the.describe() function to verify that my mean brent oil per barrel as a percentage of crude oil per cubic meter was 15.9% with a standard deviation of 0.14% meaning prices between the two sources were largely in line, and verified this by plotting the two prices trends over time, with two different scales on the y-axis. From this graph, inserted again below, we can see the two trends are almost identical, meaning it is safe to assume the data collected from Reuters, via a third party, is reliable. I finally corr() function from Pandas to calculate the correlation coefficient between the two oil prices and got a value of 0.9996, thus confirming my assumptions were correct.   
  
* Another insight I derived from my analysis is that the correlation between oil price and the daily rate shipping rate for a 310K tonne ship had changed over time. Initially, in the late 1990’s and early 200’s we can see that the two trends are closely correlated, with peaks and dips in the same years. In the early 2010’s, this trend seems to change and the appears to be an inverse relationship between oil price and daily ship rates, an unexpected result, and finally from 2015 onwards we see the two seem to be closely correlated again, but now with a time lag factored in, with daily shipping rates appearing to follow the same patter oil prices do, but a couple of years behind. This I would assume is driven by oil futures and forwards being purchased at the current days price, but for the following year or year after.

A graph of a line graph

Description automatically generated with medium confidence

* Another insight I derived from my analysis is that the smaller (74K tonne and MR Products), and thus cheaper in terms of daily rate, shipping methods appear to be becoming more popular as time goes on. As we can see from the below graph, while in the 2000’s, the larger ship (the 310 tonne) had a much higher daily rate than the two smaller ships, but as we move into the 2010’s, the gap closes, with prices being at their very closest looking at the most up-to-date data. It can be assumed that the daily price of this shipping method is closing to close to parity with the other methods as demand for the larger ship is falling due to cheaper alternatives being available.

A graph showing the price of oil

Description automatically generated

* Looking at the graph immediately above, not repasted for the sake of space, we can also derive that the daily price of the smaller ships is also closely correlated with the price of a barrel of brent oil, with the same surprising inverse relationship as the 310K tonne ship had in the early 2010’s. We can see from looking at the peaks and dips in this graph that, from roughly the year 2000 to 2010, the price of these smaller ships peaked and dipped in the same years the price of a barrel of oil did. In 2010, we see an inverse correlation develop for 4-5 years, and again similar to the larger ship, we see the daily ship prices follow the price of a barrel of oil, but a couple of years lagged. Again, in this instance I would assume it driven by the purchase of one- and two-year oil futures and forwards at “todays” price.
* A fifth insight that I have derived from this analysis is that in 2024 or 2025, we are going to see a stark increase in the daily shipping rates. As mentioned in my previous insights, there appears to be a close correlation between oil price and daily shipping rates, currently lagging a few years behind. If we examine the graph, we can see there has been a stark rise in the price of oil since 2020 and so it can be assumed daily rates will follow. It is unusual to see the time lag has increased, as we would have expected to see the increase in daily shipping rates in 2022, however if we factor in that world economies slowed in 2020 and 2021 due to the Covid-19 pandemic, and industry only really kicked back into full swing in 2022, it is understandable that the length of the time-lag would also have increased.

# References

Our World in Data. (2022, December 31). *Oil price - Crude prices since 1861 (current US$)*. Retrieved August 2023, from Our World in Data: https://ourworldindata.org/grapher/crude-oil-prices?tab=table