Advanced Data Analysis, first assignment, 2020-09-26

The file `data.txt` contains a matrix with five columns. The first three are the year, month and day for the date of each data point. The fourth column is the closing time value of the NASDAQ Composite index (in US dollars) and the fifth column is the price of a barrel crude oil (in US dollars) for the Crude Oil Prices, West Texas Intermediate.

Note that this table only includes business days, and has no values for holidays or weekends. For these exercises we will pretend that only business days exist, since on holidays and weekends trading tends to be suspended. This means you can ignore the gaps and process the data as if all days were consecutive, but you cannot assume that years have 365 days, since only days with data are counted.

When answering these questions, you should extract relevant values from the data with your Python code and not insert them manually unless the values are specified in the question. For example, if the question asks for year 2015, you can include this value in your code. But if the question asks for the first year in your data set, your answer will be penalized if you look up the value yourself to insert it manually. Note that this applies to values you use in your code. Values used in your justifications may be read from charts if they are easy to spot.

You can use any code from the lectures or exercises if you want.

When required, you must justify your answer. Please provide brief, clear and direct justifications. Invoking irrelevant factors, giving vague justifications or writing more than necessary will be penalized.

Question 1 [4 points out of 20]

Write code to print out the answers to the following questions:

1.a) How many days with data do you have for each year, for the years 2016 through 2019?

1.b) What is the average number of days with data for the years 2016 through 2019?

1.c) What is the minimum price of crude oil (fifth column) for each year from 2015 through 2020?

1.d) What was the date (year, month, day) when the crude oil price fell to the lowest value?

Question 2 [4 points out of 20]

In April of 2020 the price of a barrel of crude oil went down to negative numbers. This is not an error in your data. The break in demand due to COVID-19 was so sudden that producers had difficulty storing the excess oil being produced and had to pay to get rid of excess crude oil. So we will disregard 2020 when looking at the price trend over the years.

Also, we have incomplete data for 2015, since our time series starts only in September of this year.

So plot the data from 2016 through 2019, along with a moving average starting from mid 2016 and ending mid 2019, and chosen in a way that allows us to see the yearly change in price without the interference of seasonal variations within each year.

Add a comment to your code or a text cell in your notebook justifying your choice of moving average algorithm and the parameter or parameters you used.

Question 3 [4 points out of 20]

Starting from June 2020 (month 6 of the year) the oil price was recovering, but fell again these last few days. Plot the data from June 2020 to the end and a moving average that smooths fluctuations during each week but represents the data up to the last point and gives more weight to the latest values, in order to better reflect the trend at each time.

Add a comment to your code or a text cell in your notebook justifying your choice of moving average algorithm and the parameter or parameters you used.

Question 4 [4 points out of 20]

Due to greater oil use in the Northern hemisphere and seasonal variations due factors like heating in Winter, traveling in Summer and so on, there may be regular fluctuations of oil prices with periods of perhaps six months, or a year, or somewhere in that range.

Check the data to see if this is true using the years from 2016 through 2019. Feel free to adjust the plots and represent whatever charts you need to help you justify your answer.

Question 5 [4 points out of 20]

Some people think that oil prices can drive the stock market. Some say higher oil prices depress the stock market, others say that oil prices tend to follow the general trend of the economy. Some think oil price variations can be used to predict stock market changes in the following days, others that oil and stock prices are not related by any simple cause and effect relation so this is not true, or that oil and stock prices follow each other so closely that there is no time to use this information from one day to the next.

What do you think? Compare the NASDAQ Composite index and the crude oil price per barrel in your data set, considering the years 2016 through 2019, and justify your answer. Feel free to explore different plots and represent the data in whatever way you need to justify your answer.