

Question 1

(a) Plot a graph of the date (x-axis) versus ftse, the UK based stock index. Hover your cursor on the graph and guess the month and year when the highest value occurred.

=> The graph shows that the highest value is 7877.5, which occurred in May 2018.

b) Create 4 sub-plots one on top of the other, one for each of the four stock indices (spx, dax, ftse, and nikkei) against dates in the x-axis. By just eyeballing which of the four had the greatest dip during COVID onset in 2020?

=> The graph shows that dax Index has the highest dip (around 13000 to 9000) in 2020

(c) Using the above 4 subplots, what are the other times there is a global fall in stock markets? Can you state what events these corresponded to?

=> There are dips in the year 2003, 2009, and 2016 and the events corresponding are SARS outbreak, Global Financial Crisis Aftermath, and Chinese Stock Market Turmoil respectively

(d) Obtain a heat map of the correlations between all four indices (for the entire duration). Comment on the correlations highlighting what you expected to be correlated or uncorrelated based on the graphs. Were there any surprises?

=> From the heatmap, it seems that the SPX and DAX indices are highly correlated with each other. The FTSE also shows a high degree of correlation with both the SPX and DAX, which is not surprising given the interconnectedness of the European and American economies. The Nikkei index, while still positively correlated with the SPX, DAX, and FTSE, shows a lower correlation coefficient. This could be due to the unique economic and monetary policies of Japan

(e) Create 4 more subplots, now just using years 2005, 2006, 2007, 2008, 2009, and 2010 data. Do the four indices behave similarly? Write your thoughts about the trends.

=> All four indices behave similarly as they reach highs or lows around the same time, indicating a global market movement.

(f) Now obtain a heat map only for years 2005-2010 (both included). Which two indices were most correlated earlier for the full data and which two are most correlated now?

=> spx and dax correlated the most earlier with correlation value of 0.95 whereas now spx and ftse correlate the most with value of 0.95

Question 2

(a) Subset the data by only considering the years 2014, 2015, 2016, 2017, and 2018 for both the weather data as well as the stock index data. Which data set has NaN values? And in which columns are they?

=> London weather dataset has NaN values in columns global_radiation and snow_depth

(b) Use `df['column'].interpolate(inplace = True)` to interpolate the values of NaN as the data is already sorted by dates. State the number of rows (n) at this stage for each of the data sets and also check there are no NaNs.

=> Number of rows in Weather dataset and Stock dataset are 15341 and 7255 respectively

(c) Use only the date and 'ftse' columns from the stock data, and merge those columns with the London weather data. Use the date field as the merge key. Use all the rows of the weather data. How many NaN rows are in the resulting set?

=> Number of NaN rows in the london weather dataset after merging are 8284

(d) The stock market does not have any data published on holidays. Fill those NaN using interpolate. Also drop the column 'Date' as it is the same as 'date'. How many rows of NaN are in the merged dataset now? Also, how many columns are in the merged set now?

=> Number of NaN rows in the merged dataset after interpolation are 5485 and Number of columns in the merged dataset are 11

(e) Obtain a heat map of the correlations between all the numerical columns but only for a subset of merged data when snow depth is greater than zero. So looks like the closing index value is dependent on the weather that day provided there was some snow depth! Which variables is 'ftse' most and least (i.e. most negative) correlated?

=> Variable most correlated with 'ftse' is date and Variable least correlated (most negatively) with 'ftse' is sunshine