CIS 662 Spring 2024 Introduction to Machine Learning and Algorithms
Assignment 1 – Part 2 – Take Home
Name: Rahul Manjunath Pudurkar SUID: 791804558 Email – rpudurka@syr.edu

Question 1:

a) The highest value of the **FTSE index** within this dataset occurred in **May 2018**, when the index reached a closing price of **7877.5**.

b) By looking at the subplots we can estimate that during the onset of Covid in 2020 **ftse** stock had the **greatest dip** in its value within all the stock indices.

c) There was a notable decline in stock market indices during the **"2008 stock market crash."** The year **2000**, which saw the **"dot-com crash,"** shows another slight decline.

d) The **UK's 'ftse'** index and the **German 'dax'** index have a strong correlation of **0.87**, while the US's spx index and the UK's 'ftse' index have a strong correlation of **0.77**, as predicted. Additionally, the German **"dax"** and **"Japanese "nikkei"** have a lower correlation, as predicted, with a value of **0.55**. The weak correlation between the Japanese "nikkei" and UK "ftse" indexes (correlation coefficient of **0.37**), was one unexpected finding.

e) Between 2005 and 2010, the following changes are noted.
- The positive correlation between spx v/s nikkei (0.63->0.89) and spx v/x ftse (0.78->0.95) has somewhat increased.
- The strength of the positive correlation between dax and spx has somewhat decreased (0.95->0.78).
- The positive correlation between ftse and nikkei increases significantly; it changes from being moderately associated to strongly correlated (0.37->0.84).
- The strength of the positive association between ftse v/s spx has also somewhat increased (0.78 -> 0.95).

f) With a correlation coefficient of 0.95, the US "spx" and German "dax" indices were the most associated in the earlier heatmap that featured all the dates. When the data is restricted to only include dates from 2005 to 2010, the US "spx" and UK "ftse" indices have the highest correlation, with a value of 0.95.

Question 2:

a) The London weather dataset has Nan values. These Nan values are present in the **global_radiation(6 rows)** and **snow_depth(2 rows)** columns.

b) There are **1826 rows** in the filtered **London weather dataset** and **1304 rows** in the filtered **stock index dataset**, respectively. Additionally, following interpolation, none of the dataset's columns have any NaN values.

c) There are **522** NaN rows in the merged dataset.

d) There are **no NaN** rows once the **"ftse"** column is interpolated and the "Date" column is dropped as it is repeated. Additionally, there are **11 rows** in the final dataset.

e) The ftse index is most correlated to **precipitation** (0.94) and least correlated to **snow depth** (-0.819).