Question 2

a.

* In the past 2 years, the WHO’s database has nearly doubled. That could mean for example lack of data for the same cities for the past years
  + Last updated in 2018, covers data up to year 2016.
  + Covers 4300 human settlements, in 108 countries.
  + One source is reporting by country’s official. Possible misrepresentation, human error or lack of calibration of pollution sensors.
* Sources outside London can affect air pollution in London.
  + Between 45-55% originates outside of London, and can be of human or natural origin (e.g. Saharan dust) [1]

Word count: 93

References:

[1] London.gov.uk. (2012). *Air Quality Information for Public Health Professionals –London Borough of Southwark*. [online] Available at: https://www.london.gov.uk/sites/default/files/air\_quality\_for\_public\_health\_professionals\_lb\_southwark.pdf [Accessed 1 Mar. 2020].

b.

The filename of the notebook relevant to this part of question 2 is “TMA02\_Question2b.ipynb”

c.

Report investigating patterns in Air Quality data in suburb of London.

* Aims and objectives

To explore whether levels of different types of pollutants are interrelated with each other.   
To measure whether different days of the week (e.g. Monday, Tuesday etc.) affect pollutant levels.

* Background

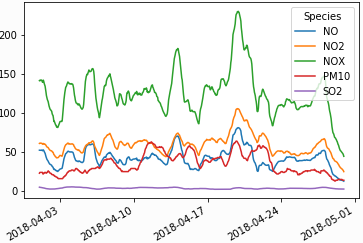
Air pollution in London is a growing problem, which affects populations’ health and healthy life expectancy causing in 2008 estimated 4300 premature deaths [2]. This report will explore patterns in Air Quality in London Borough of Lambeth.

* Sources of data

Data was collected for publicly available sources at the website of the London Air Quality Network (LAQN) and comes with permission for non-commercial use [5].   
  
Londonair is provided by the Environmental Research Group of King's College London [4].   
  
The data comes from the Lambeth Bondway Interchange station, with 2 metres sampling height, and 3 metres distance to road [3] with averaging period of 1 hour.   
  
The data set covers the month of April 2018.

* Analysis pipeline

Data was imported into Panda’s (software library written for the Python programming language) dataframe and then processed using SciPy library used for scientific computing.   
  
The data values were cleaned of any non-numeric values.  
  
Spearman's rank correlation coefficient was measured for pollutants grouped by mean (average) day of the week figures, as well as by each recording period  
  
Chi-Square test of independence was performed on data grouped by mean (average) day of the week figures, as well as by each recording period.



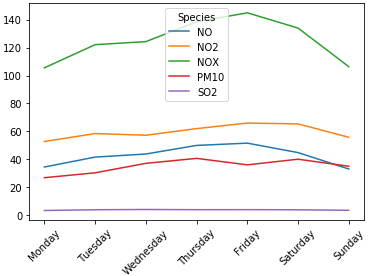


Figure 2: plot corresponding to measure of pollutants for grouped by mean (average) day of the week figures.

Figure 1: plot corresponding to measure of pollutants for each measuring period.

* Findings

Spearman's rank correlation coefficient resulted in positive relationship between variables, both when grouped by average recording period and when grouped by mean figure by day of the week for all types of pollutants.  
Findings were more correlated when the pollutant values were grouped by day of the week, although they were more significant (lower p value in all cases except in 'PM10' particles vs Sulphur Dioxide) when grouped by each recording time and date.  
  
Chi-Square test of independence resulted in p= 0.00 when grouped by average recording period.  
Chi-Square test of independence resulted in p= 0.99 when grouped by mean figures by day of the week.

* Conclusions

Between 45-55% of air pollutants in London originates outside of London and can be of human or natural origin (e.g. Saharan dust, forest fires etc.) [1]. Lack of calibration/standardisation of measuring equipment and potential human errors should be considered.

Because of low p value, we can conclude that pollutant levels are dependent when grouped by each recording period/hour.   
  
From high p value we can conclude that variables are independent, i.e. not linked together when grouped by mean figures by day of the week. We can say that day of the week does not affect each pollutant levels separately, which is confirmed by higher, on average, correlation (SpearmanR) between pollutants when measured by day of the week.

It is important to remember that data from only one month and one measuring station was used, and could only be thought of as a sample, not a population.  
More analysis could be done to find what how day of the week affects pollutant levels in different parts of the city and during different seasons.

* References

[1] London.gov.uk. (2012). *Air Quality Information for Public Health Professionals –London Borough of Southwark*. [online] Available at: https://www.london.gov.uk/sites/default/files/air\_quality\_for\_public\_health\_professionals\_lb\_southwark.pdf [Accessed 1 Mar. 2020].

[2] London City Hall. (n.d.). *Health and exposure to pollution*. [online] Available at: https://www.london.gov.uk/what-we-do/environment/pollution-and-air-quality/health-and-exposure-pollution [Accessed 3 Mar. 2020].

[3] London City Hall. (n.d.). [online] *Station Details* Available at: https://www.londonair.org.uk/london/asp/publicdetails.asp?site=LB5 [Accessed 4 Mar. 2020].

[4] Environmental Research Group, K. (2020). London Air Quality Network. [online] Londonair.org.uk. Available at: http://www.londonair.org.uk/LondonAir/General/about.aspx [Accessed 5 Mar. 2020].

[5] Londonair.org.uk. (2020). *London Air Quality Network :: Welcome to the London Air Quality Network » Statistics Maps*. [online] Available at: http://www.londonair.org.uk/london/asp/copyright.asp [Accessed 5 Mar. 2020].

Word count: 550 (not including references)