This data summary aims to provide insights about the air quality levels in 2020 across 4 boroughs in London - Barking and Dagenham, Hackney, Lewisham, and Sutton. The air quality levels are assessed by the presence of 3 pollutants namely Nitrogen Dioxide (NO2), Nitrogen Oxide (NOX), and Nitric Oxide (NO). High levels of these gases could be indicative of increased use of vehicles, increased use of fertilizers (during agriculture), or an increased in industrial activities. The primary data set used for this report, air_quality_4_boroughs_2020.csv, consists of 35136 rows and 12 columns. Additionally, two other data sets, ethic-group-by-borough_2020.csv and hourly_wage_by_borough_2020.csv, have been used to analyze trends The variables and corresponding data types from the original data set are as follows:

Site - object

NO2 - float

NOX - float

NO - float

ReadingDateTime - object

Since_start_of_year - float

NO2_rolling_daily_mean - float

NOX_rolling_daily_mean - float

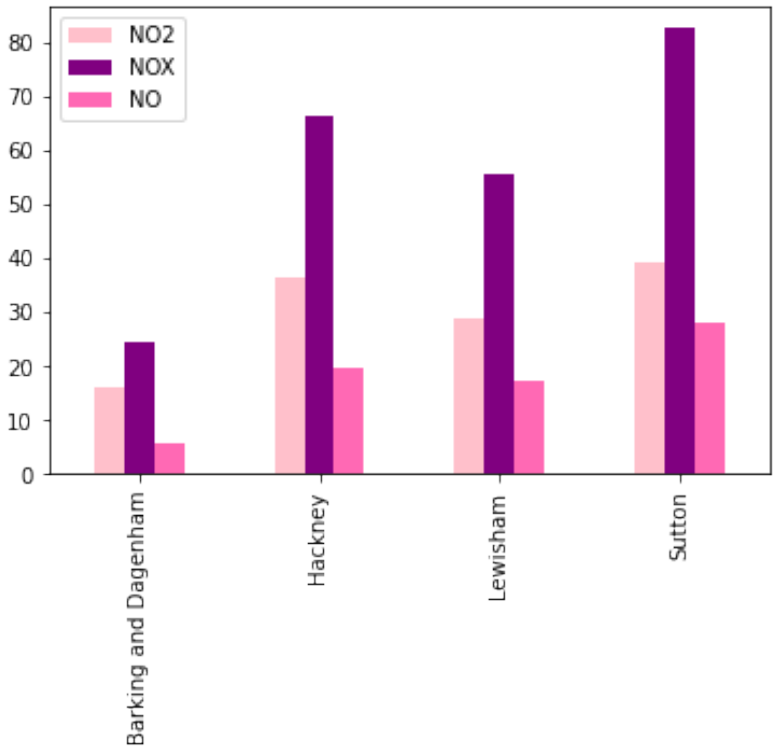
NO_rolling_daily_mean - float

NOX_rolling_daily_std - float

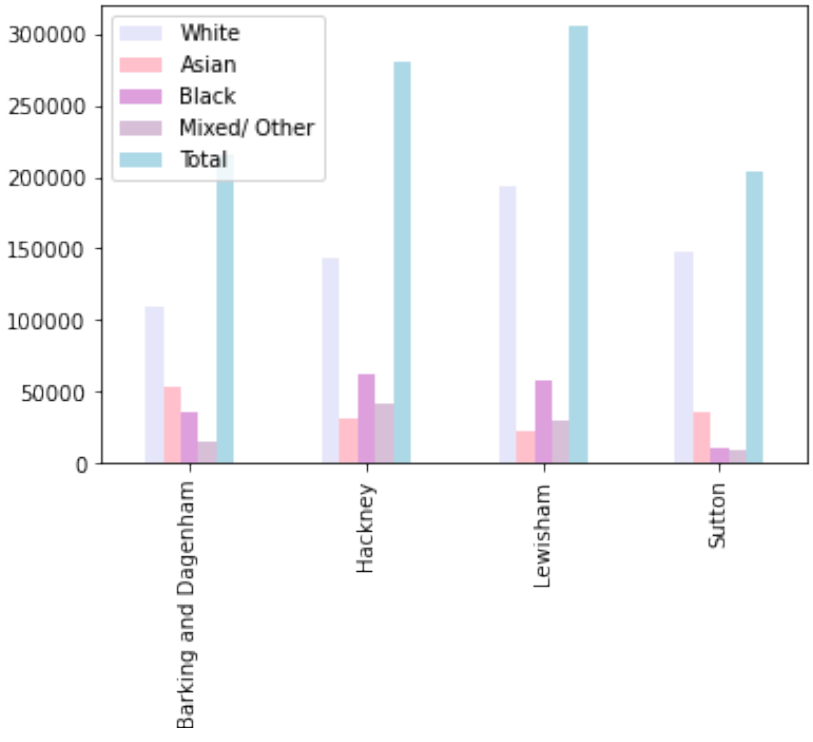
NO2_rolling_daily_std - float

NO_rolling_daily_std - float

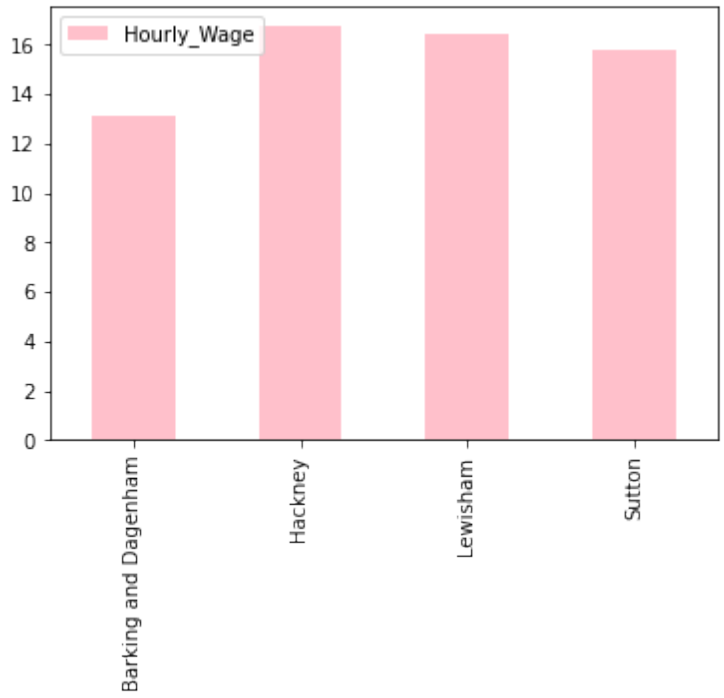
The mean level of NO2, NOX, and NO across the 4 boroughs are 30.07, 56.88, 17.48 respectively. The highest recorded level of NO2 at 191.6 was in Sutton on 07/08/2020 while the lowest recorded level of NO2 at -3.8 was at Barking and Dagenham on 12/03/2020. The highest recorded level of NOX at 961.8 was in Sutton on 21/01/2020 while the lowest recorded level of NOX at -1.0 was at Hackney on 04/06/2020. The highest recorded level of NO at 510.1 was in Sutton on 21/01/2020 while the lowest recorded level of NO at -1.7 was at Hackney on 16/09/2020.



From the image above, it can be seen that Sutton has the most pollutants in the air with an average of 39.4 NO2, 82.6 NOX, and 28.2 NO, suggesting that it has the worst air quality of the 4 boroughs. On the other hand, Barking and Dagenham has the least pollutants in the air with an average of 16.0 NO2, 24.6 NOX, and 5.6 NO, making it the borough with the best air quality of the 4. Lewisham has slightly less pollutants in the air than Hackney, making it the borough with the second best air quality.



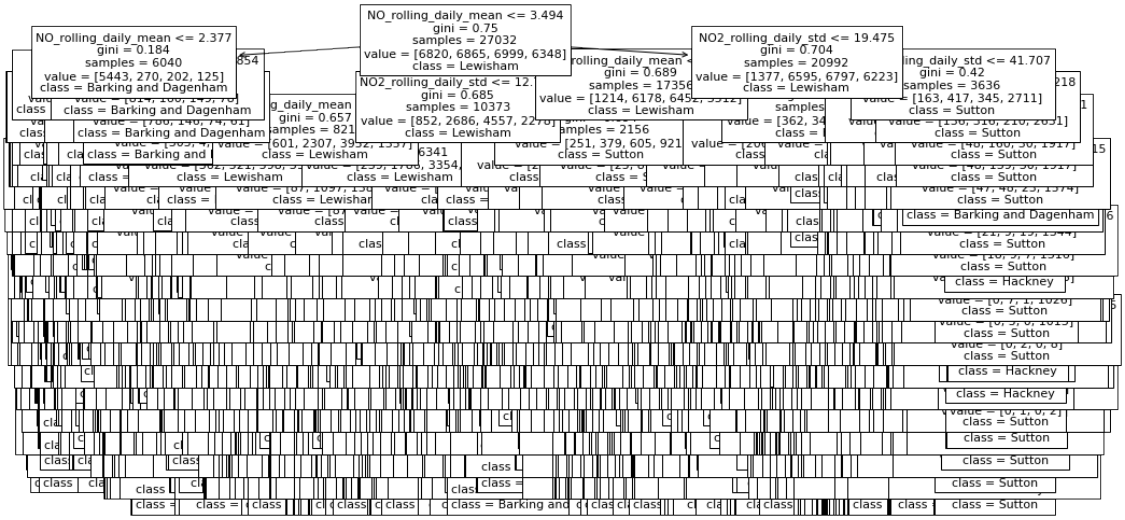
In order to see how population plays a role in the air quality level, the population sizes and ethnic makeup of the population in each of the 4 boroughs were plotted. As seen from the graph above, Sutton has the smallest population at 204000 while Lewisham has the largest population at 305000. Hackley has the second biggest population at 281000 and Barking and Dagenham has the second smallest population at 215000. From the graphs, it can be observed and assumed that the population size does not significantly affect air quality levels since there is no observable trend.



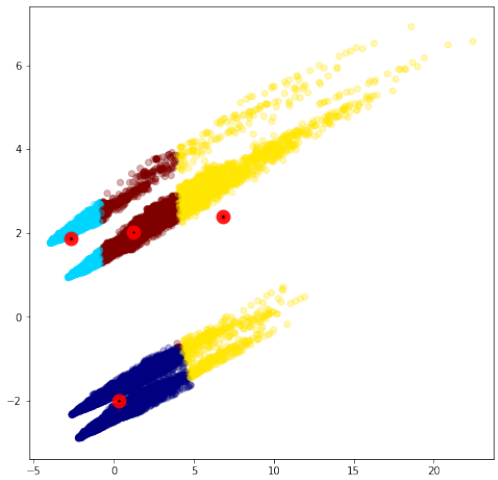
In order to see how the average wage in a borough plays a role in the air quality level, the average hourly wage in each of the 4 boroughs were plotted. As seen from the graph above, Barking and Dagenham has the lowest average wage at 13.1/h, while Hackney has the highest average wage at 16.74/h. The average wages at Lewisham and Sutton are marginally less than at Hackney at 16.46/h and 15.8/h respectively. The significantly low average wage and the significantly low pollution levels in Barking and Dagenham suggest that there indeed may be a relationship between average wage and pollution levels. This claim is further supported by the slight positive correlation between hourly wage and level of pollutants as seen in the correlation matrix (in the Notebook). This trend could be due to the polluting/non-polluting natures of the predominant industries active in each borough. For example, the dominant industry in Barking and Dagenham could be more manual/un-skilled and may require less machinery, hence explaining the low average wage and proportionately low levels of pollution.

As seen from the correlation matrix, many of the variables have correlations with each other, allowing us to build a model using one or more variables to predict another or demonstrate a relationship with two or more variables. When a multiple variable regression model was trained with ‘NO2’, ‘NOX’, ‘NO’, ‘NO2_rolling_daily_mean’,'NOX_rolling_daily_mean','NO_rolling_daily_mean','NOX_rolling_daily_std','NO2_rolling_daily_std', and ‘NO_rolling_daily_std' as variables, it was able to predict hourly wage. However, owing to the fact that correlation between these variables and hourly wage were not very strong, the R2 score was low at 0.22.

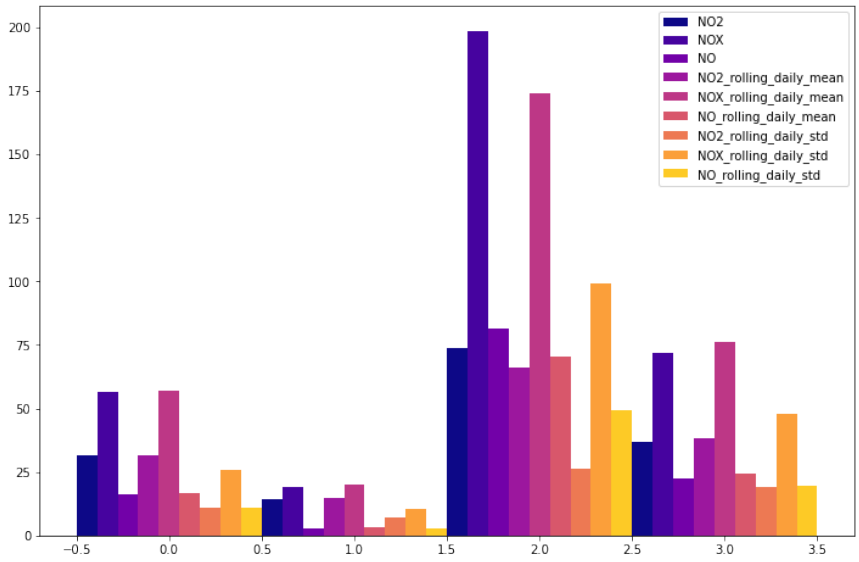
The data can also be used to build a model that predicts a categorical class for a new example. For instance, a model with a 79.5% accuracy was trained using ‘NO2’, ‘NOX’, ‘NO’, ‘NO2_rolling_daily_mean’, ‘NOX_rolling_daily_mean’, ‘NO_rolling_daily_mean','NOX_rolling_daily_std','NO2_rolling_daily_std', and ‘NO_rolling_daily_std' as variables, to predict the y variable ‘Site’,using a max depth of 20. This information is represented in the form of a decision tree classifier as shown below:



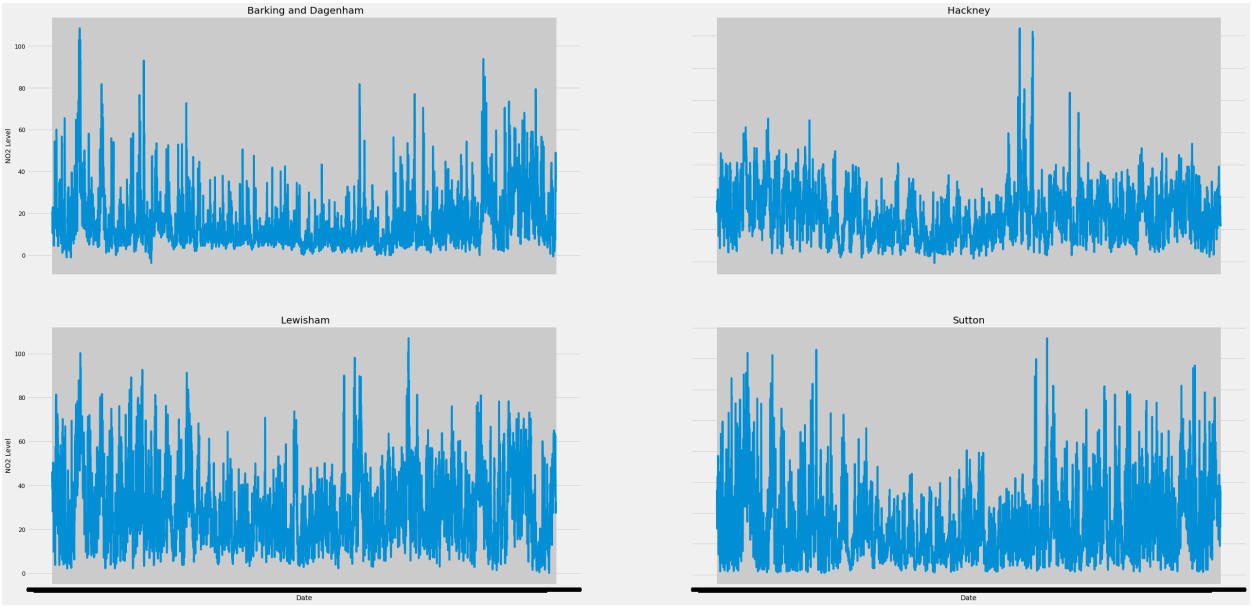
The data can be clustered into recognize groups without the need for labels as seen below. The clusters evidently show that there are 4 unique clusters representing each of the 4 boroughs.



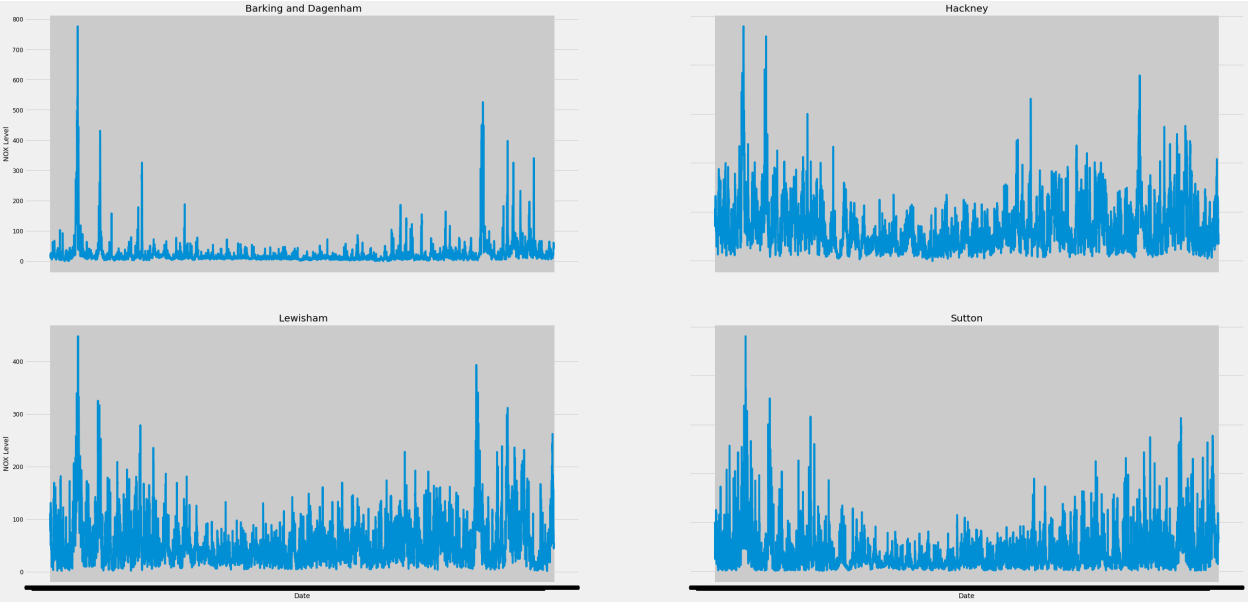
This distinct clustering can be explained by the graph below whereby the pollutant levels, especially NOX, are in very different scales for each of the boroughs.



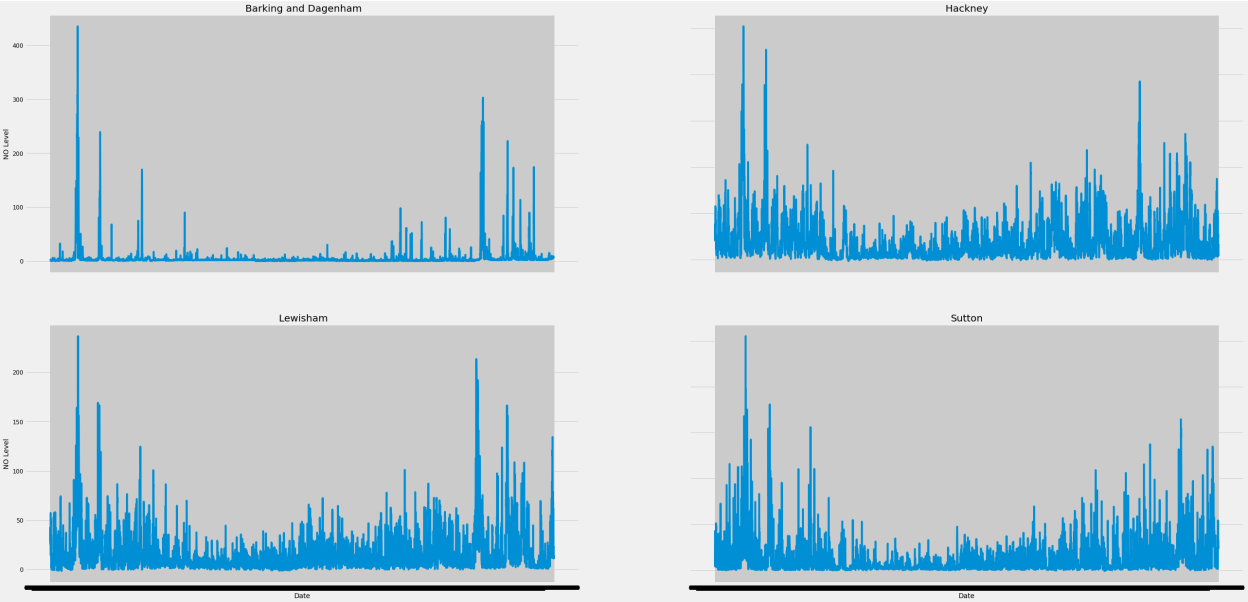
As seen in the plots below, NO2, NOX, and NO levels fluctuate in each borough fluctuates through the year. The NO2 level is high in the beginning and end of the year in Barking and Dagenham and Sutton but is generally low throughout the middle of the year. However, in Hackney, the NO2 levels are generally low through the start and end of the year, but spikes up for a short period during the middle of the year. In Lewisham, NO2 levels are generally consistent throughout the year with occasional spikes on some days. In general, Sutton has the highest NO2 levels throughout the year while Barking and Dagenham has the lowest .



NOX levels are extremely low throughout the year in Barking and Dagenham except for an extreme spike in the start of the year and a few moderates spikes at the end of the year. The NOX levels in Hackney, though higher than in Barking and Dagenham, is generally low throughout the year except for a few spikes in the start, middle, and end of the year. Lewisham and Sutton share a similar trend in NOX levels whereby the levels are generally low throughout the year but spikes in the start and end of the year. On average, Sutton has the highest NO2 levels throughout the year while Barking and Dagenham has the lowest.



The general trend of NO levels in all boroughs are similar whereby there are spikes in the start and end of the year and the levels are almost consistent throughout the rest of the year. Barking and Dagenham has the lowest NO levels through the year while Sutton has the highest.



The constantly high average levels of NO2, NOX, and NO in Sutton could lead to conclude that despite the small population, they might have the greatest use of vehicles per unit of population or industrial activity, causing them to have the worst air quality of the 4 boroughs. The opposite could be said in the case of Barking and Dagenham, where the use of vehicles per unit population or level of industrial activity may be very low, causing them to have the best air quality of the 4 boroughs.