**IBM Data Science Capstone Project**

**on Coursera**

**October 2020**

**By**

**Olumide Ojo-Oratokhai**

**INTRODUCTION**

**Problem Description:**

There is a big problem of plastics waste ending up in the unintended places or environment. The idea of this project is to explore and understand the locations of some key touch points, facilities and stakeholders in the plastic supply-chain (like plastic manufacturers, supermarkets, plastic recyclers) who can influence and help tackle the issue. The project will compare these clusters within a developed city like London and compare to those in a rapidly developing city like Lagos, Nigeria. The results will present locations of facilities and resources that stakeholders (like consumers, entrepreneurs, product designers and waste management policy makers) will find helpful when they are trying to dispose plastic waste, source recycled materials to design products and where to establish new waste handling facilities.

**Background:**

Densely populated cities like London and Lagos are big generators of waste in general and waste management is a key issue to keep atop of. Time poor urban dwellers are heavily reliant on convenient groceries and products that are mostly packaged in single-use plastics to ensure the products reach the consumer safely and last till the desired expiration. This dependence on single-use plastics has led to decades of piling plastic wastes that are not recycled. Fortunately, there have been increasing awareness of the damages that plastic wastes in particular cause to our environment and wildlife, especially when they are not recycled and end up in unintended/undesirable locations like water bodies and landfills. This is thanks to documentaries like Blue Planet II by Sir David Attenborough in 2017 and organisations like The Ellen MacArthur Foundation who highlight the dangerous impacts of unmanaged plastic waste that will eventually enter the human food chain.

There is a challenging balance to strike because our modern global lifestyles currently rely a lot on plastics in many industries from automotive, construction to food packaging to name a few. Solving such a big problems requires the involvement of many stakeholders in the supply-chain of plastics packaging ranging from resin manufacturers, packaging manufacturers, brand manufacturers, professional packaging consultants, distributors, supermarkets and consumers to waste management bodies. This project will aim to find and overlay the locations of supermarkets, plastic packaging manufacturers and plastic recycling points over a map of a developed city where there is some 'perceived' good level of waste management, and compare to a developing city like Lagos with challenging waste management issues. This comparison may bring about insights that can influence consumer disposal behaviour and other policy/decision makers to decide if, how and where to invest resources that can help to substantially reduce the leakage of plastic packaging into undesired environments.

**DATA**

**Data Description:**

The sets of data to be used in the project are:

1. Location coordinates of supermarkets and hyper-markets within 30km radius of a central neighbourhood/point in London and Lagos. (Tower of London, and Lagos state Polytechnic, Isolo Campus). (Data source: Foursquare API).

2. Location coordinates of plastics packaging manufacturers within 30km radius of a central neighbourhood/point in London and Lagos. (Tower of London, and Lagos state Polytechnic, Isolo Campus). (Data source: Foursquare API).

3. Location coordinates of plastic recycling points/recyclers within 30km radius of Tower of London and Lagos Polytechnic.   
(Data source: Foursquare API).

4. Location coordinates of packaging professional consultancies within 30km radius of Tower of London and Lagos Polytechnic. (Data source: Foursquare API).

5. The population of London. (Data source: <https://populationstat.com/united-kingdom/london>)

6. The population of Lagos. (Data source: <https://worldpopulationreview.com/world-cities/lagos-population>)

7. Tonnage of plastic wastes generated and recycled proportions in London and Lagos (Latest historical of 2018). <https://www.circularonline.co.uk/news/londoners-plastic-recycling-could-build-the-shard-5-times/>

<https://businessday.ng/businessday-investigation/article/plastic-waste-chokes-lagos-despite-potential-billion-naira-recycling-industry/#:~:text=In%20essence%2C%20about%202%2C250%20tons,a%20USD%20250%20million%20industry.>

<https://ourworldindata.org/plastic-pollution#mismanaged-plastic-waste>

**How the data will be used:**

Foursquare API will be used to explore the businesses around a 30km radius of a 'central' city location in London and Lagos. The businesses will be explored and their categories will be used to identify the type of business - whether supermarket, packaging manufacturer or plastic recycler. A k-means cluster analysis will be conducted to see how the businesses are grouped. The results of clusters will be overlaid on a map of the cities. Then a comparison of the clusters in London and Lagos will be made, (whilst taking into account the city population and per capita plastic waste generation) to identify any similarities or differences that can provide useful insights to stakeholders towards reducing the escape of plastic waste into unintended environment.

**Data challenges and limitations:**

The above are an ideal description of data that was aimed to be used for the project. However, in reality quite a few of the data were either of poor quality and sometimes unreliable or even non-existent. Examples include the data and results from API calls for Lagos. There were limited or even unlisted venue results for categories like recycling facilities and waste disposal sites and factories, despite trying many different search queries and category ids. The discrepancy was obvious when google searches produced more results than the API calls. There were slightly better outputs for supermarket searches. Due to some of the challenges above, some of the desired analyses were not carried out adequately. <https://africacheck.org/reports/what-a-waste-fact-checking-four-claims-about-nigerias-garbage-problem/> corroborates the challenges and discrepancies in data about Lagos.

**METHODOLOGY**

This section describes the data analysis methods and processes used in this project.

**Data Acquisition and Preparation:**

*Foursquare API Calls*

Four square API calls were made using previously set up credentials to identify central locations in Lagos and London around which searches were made for different venue categories like waste and recycling sites, supermarkets and factories/industrial estates - see code lines 3 - 35 for Lagos and 36 - 118 for London. Due to limitation of 50 results per call and the multitude of supermarket venues in London exceeding that limit by far, multiple calls were made to focus/retrieve data for just a small area close to central London instead of the planned 30km radius from central London.



Example of API search call for recycling facilities around a central location in Lagos.

*Webscraping with Beautiful Soup Package*

After a web search, a wikipage was found with data on Local Authorities (LA) in Lagos and population breakdown by LA. Although the source of data was not available on the wikipage, and the reliability of the data is questionable, it was about the only page that had such information easily available.

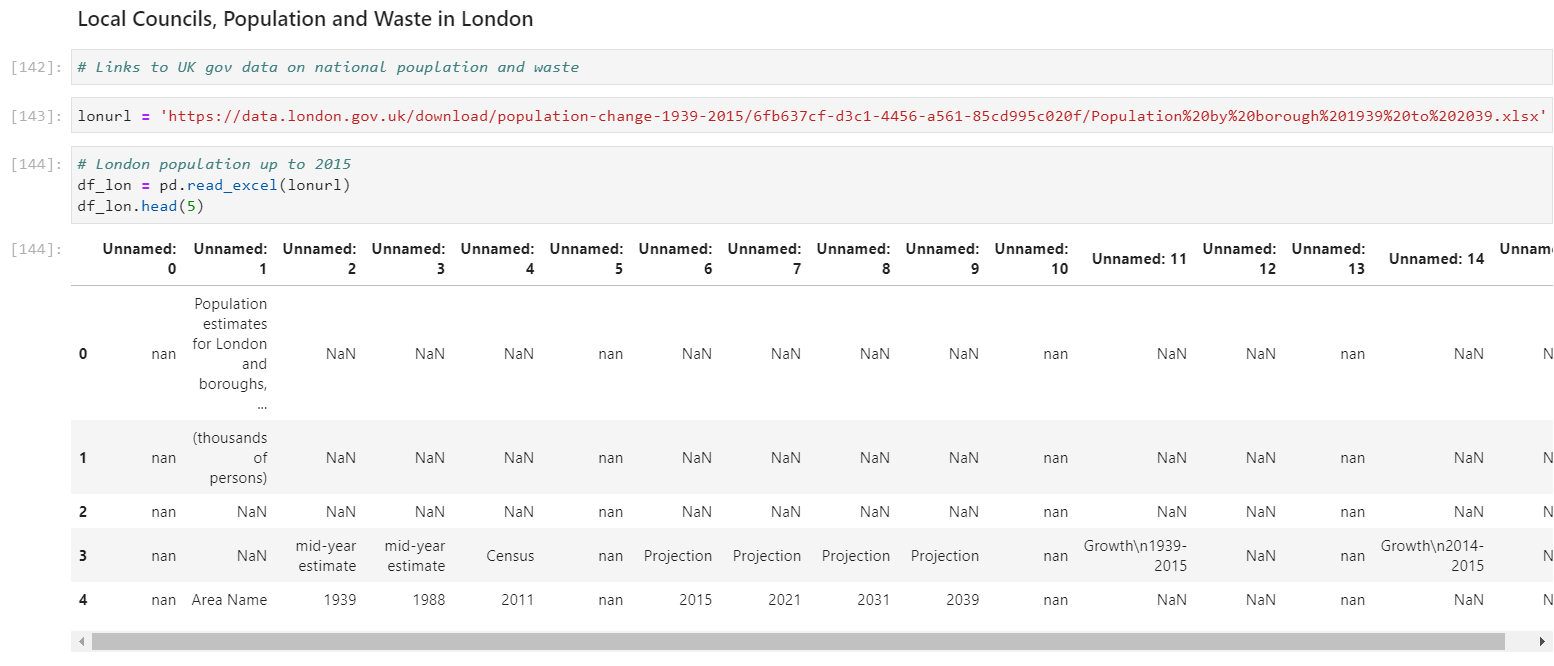
Beautiful soup package was used to scrape the data from the website into a usable dataframe - see code lines 119 -129.

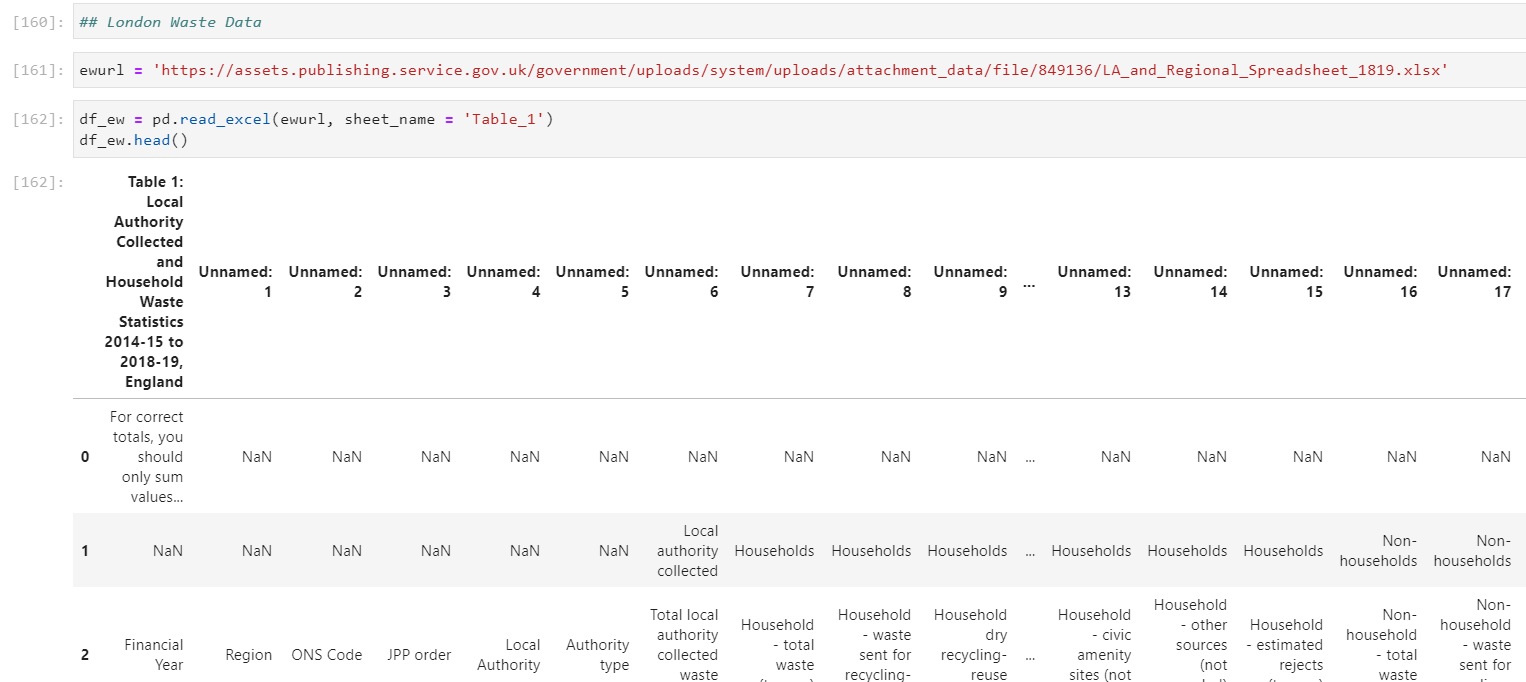


Excerpts of code for web scraping with Beautiful Soup package.

*Reading Excel Files from Urls*

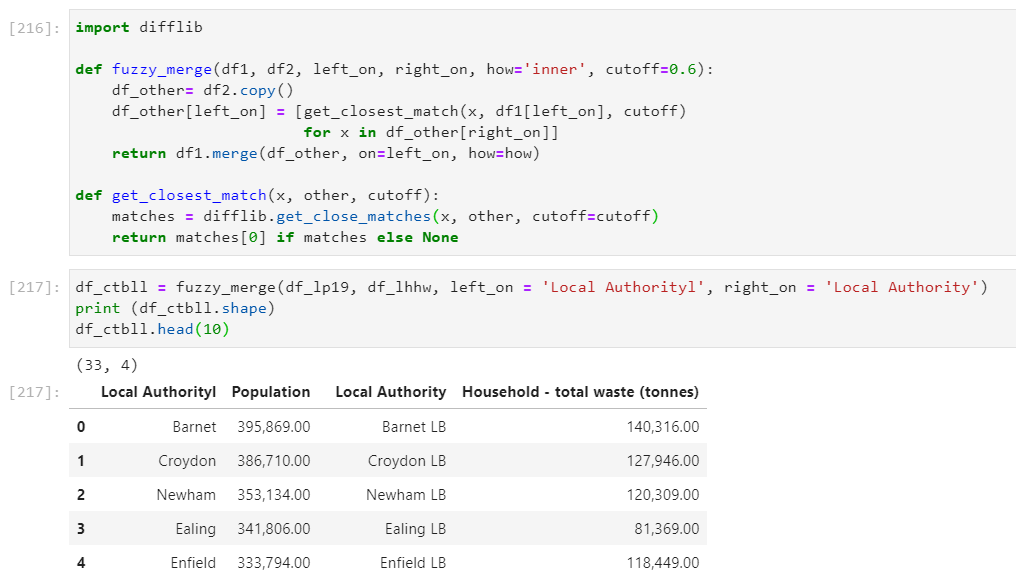
Data on London population and England-wide population and waste by Local Authorities were available from Office of National Statistics (ONS). A few different websites and files were read directly into the notebook and the data was cleaned up to focus on data for year 2019 - see code lines 142 – 194. Another more useful population data was sourced and used in code lines 195 – 214.





Codes to read Population and Waste data by Local Authority.

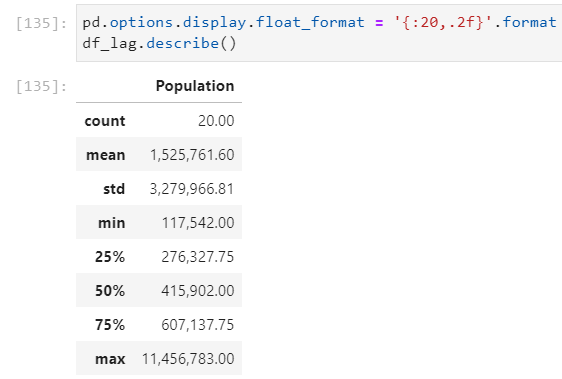
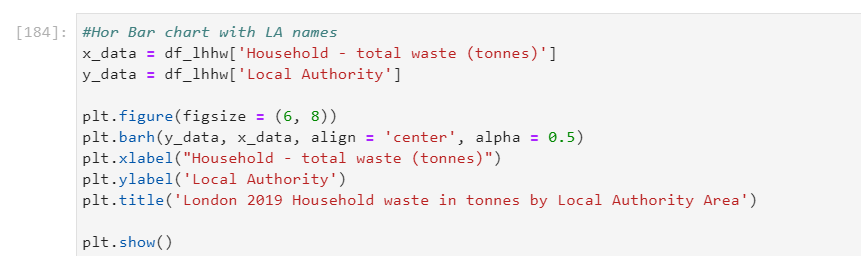
Fuzzy merge package (code lines 216 - 233) was used to be able to merge the data from different sources based on common columns that did not have exact matching Local Authority naming convention.



Fuzzy merge codes to combine tables

**Basic Data Summary and Statistics:**

After cleaning and preparing all the data extracted from the web for Lagos and London venues, populations and waste generated, the subsequent dataframes were analysed with .describe() to obtain summary statistics and highlight Local Authorities with minimum and maximum values accordingly. See code lines 135-138 for Lagos. See code lines 156 – 159, 179, 191, 204, 210, 222 and 230 for London. Summary charts were also provided for Lagos data – code line 141 and for London data in code lines 184 and 215.

Lagos Population Summary and England and Household Waste by Local Authority Chart Codes

**Data Visualisation:**

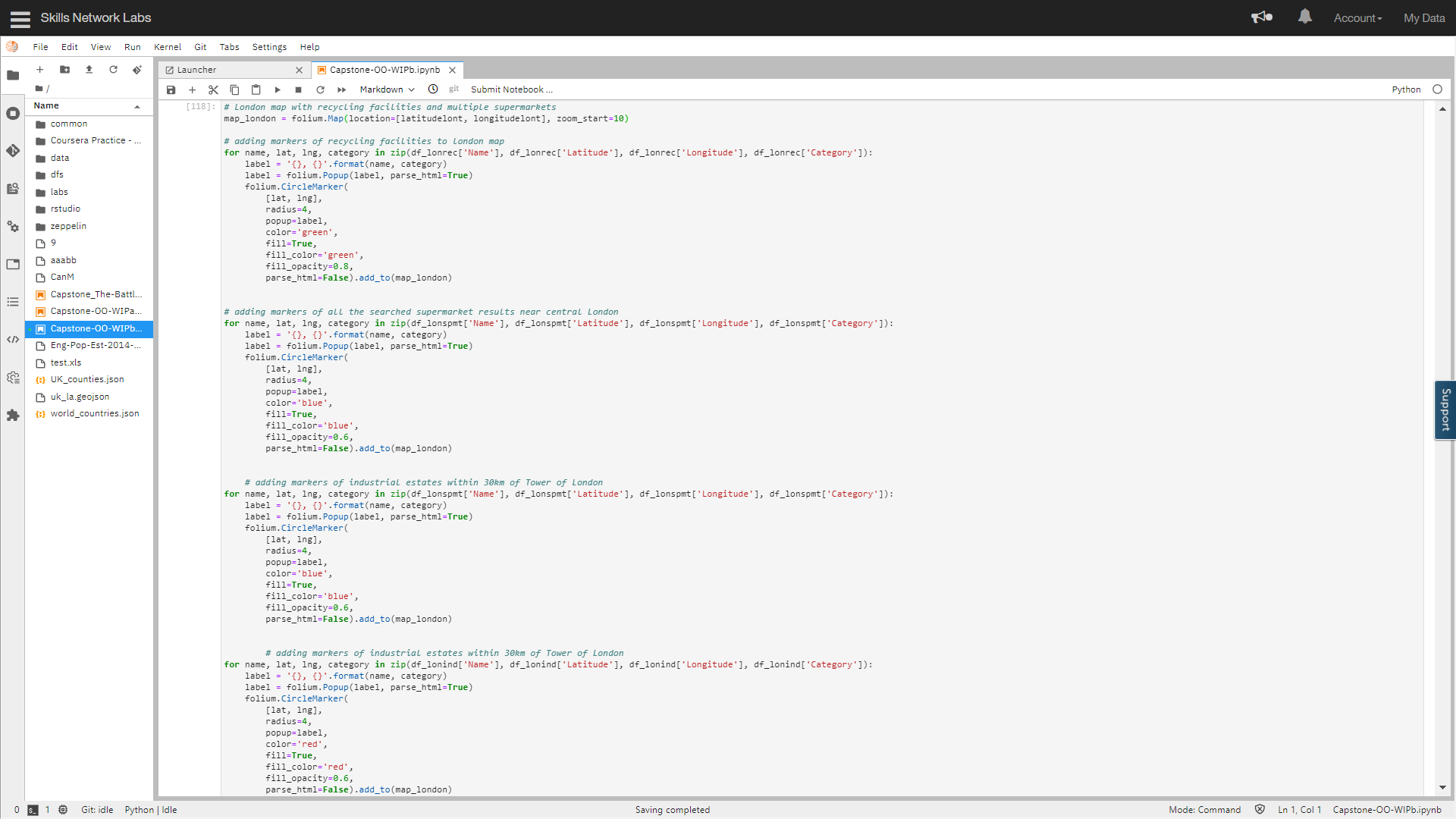
Folium maps were used to show the locations of various venues of relevance in both Lagos and London. This was used to visualise locations of waste recycling facilities, supermarkets and

factories close to the centres of Lagos and London. See code lines 14, 24, 35 for Lagos and code lines 47, 56, 109 and 118 for London.

Scatter plots were used to illustrate London data relationship between waste and population. See code lines 236-239.



Folium map codes for Lagos Venues – Recycling facilities, Supermarkets and Industrial Estates.



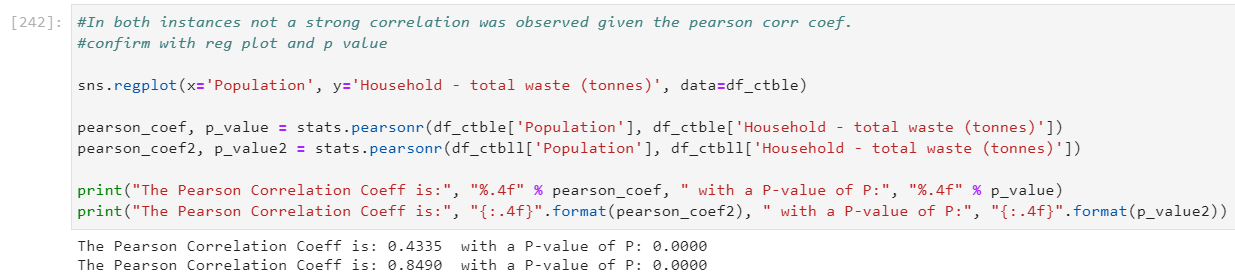
Folium map codes for London Venues – Recycling facilities, Supermarkets and Industrial Estates.

Attempts were made to create choropleth maps of waste and population across England Local Authorities. This attempt was unsuccessful as the right key\_on for the geojson file was not correctly identified and the rendered map was therefore blank - see code lines 274 – 278.

**Regression Analysis:**

Given the low volume and quality of data available for Lagos, (for instance no data on waste by Local Authority for Lagos was available) it was decided that the regression analysis focus on London Local Authorities waste and population for which the data obtained was most reliable. To have a comparison, the regression analysis was also conducted for all England Local Authorities alongside London Authorities.

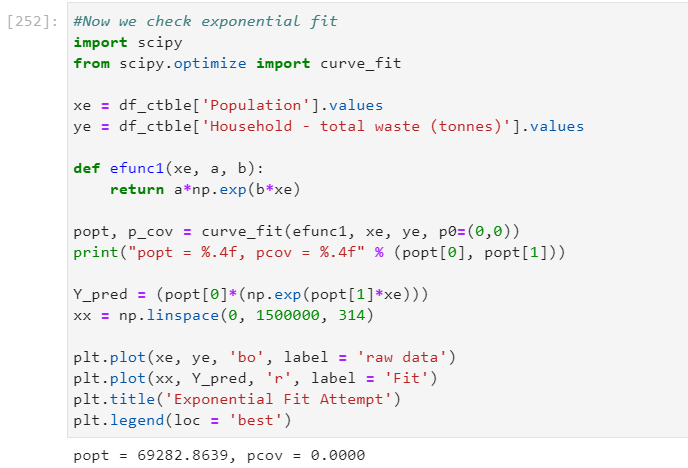
Linear correlation was explored in code lines 240-243. There was a reasonable fit for the London data but less so for the England data. Polynomial fits were attempted in code lines 244 - 251. Exponential fit was attempted in code lines 252 – 254 using the original data and normalised data to see if it’s a better fit for the England data. Finally, Logistic/Sigmoidal fits were attempted in code lines 255 - 260.



Linear fit attempt



Polynomial fit attempt



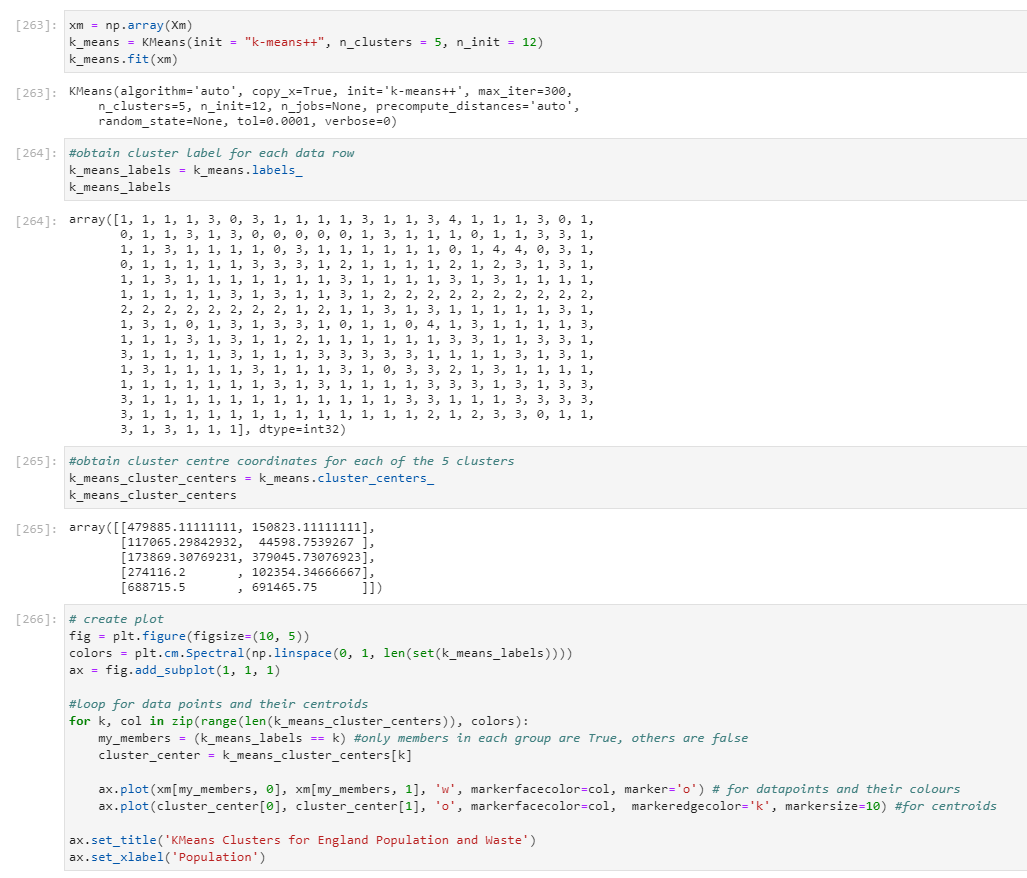
Exponential fit attempt



Sigmoidal fit attempt

**Cluster Analysis:**

K-means was used to conduct cluster analysis and group the England and London Local Authorities based on scatter plot of population vs household waste. See code lines 262-267 for cluster analysis based on all England Local Authorities populations and household waste. See code lines 268 – 273 for cluster analysis based on London Local Authorities populations and household waste.



Codes for k-means cluster analysis – England LAs data



Codes for k-means cluster analysis – London LAs data

In summary, Beautiful Soup package was used to acquire some of the data, A lot of data wrangling was carried out to tidy the dataframes and make them ready for analysis. Visualisation tools like Folium maps were used to visualise locations of venues like recycling facilities and supermarkets on Lagos and London maps. Scatter plots and horizontal bar charts were used to illustrate alongside some of the descriptive statistics of the acquired data. Attempts were made to fit linear, polynomial, exponential and logarithmic models to the data. Finally, k-means clustering was applied to the England and London Local Authorities population and waste data to highlight distinct groups within the dataset.

**RESULTS AND DISCUSSIONS**

This section entails the results of the analysis conducted using the tools described in the previous section, along with some discussions and interpretations of those results.

**Lagos Recycling Facilities, Supermarkets and Factories:**

There was limited information for the amount of waste generated by Lagos even from official sources, and this is was not possible to obtain such information at a Local Authority level. Whilst some data exist for total waste for some cities in Nigeria, none had the data at a Local Authority level. Foursquare API calls show only two waste recycling (Green dots) facilities. Although other sources like google revealed up to 19 facilities. There are 24 supermarkets (Blue dots) and 47 Factories/Industrial Estates (red dots) within 30km radius of Laspotech, Isolo, Lagos.

It would have been expected that all the search results would be much higher that what is captured in the Foursquare API database, so these results may not be as useful as expected.



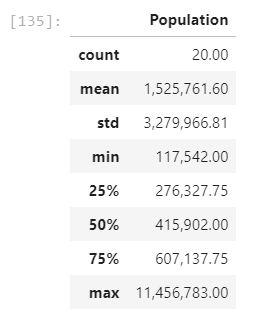
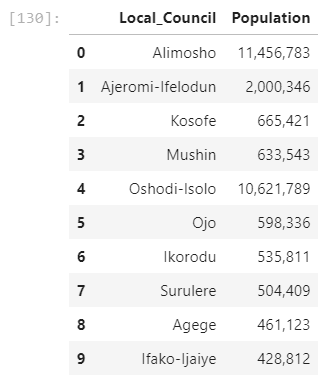
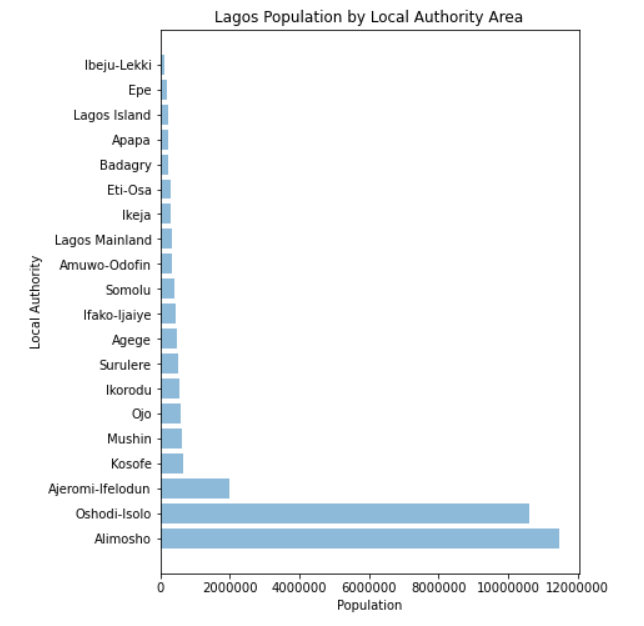
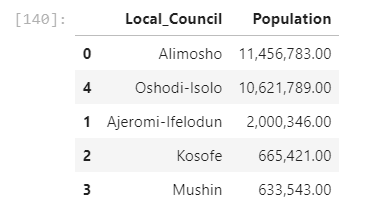
Lagos Folium map showing recycling centres (green), supermarkets (blue) and ind. estates (red)



Lagos google map showing waste and recycling facilities in Lagos.

**Lagos Population by Local Authority:**

The summary and other results tables below show Lagos having 20 Local Authorities with a total of 30.5mio people. This is an average of 1.5mio people per LA with Alimosho having the largest population of 11.5mio people. Oshodi-Isolo is next with over 10.6mio people and Ibeju-Lekki with the smallest population of 117.5k people. No sources were cited for this data on Wikipedia page, so its reliability is unknown. This again makes it difficult to use such data for reliable analysis.

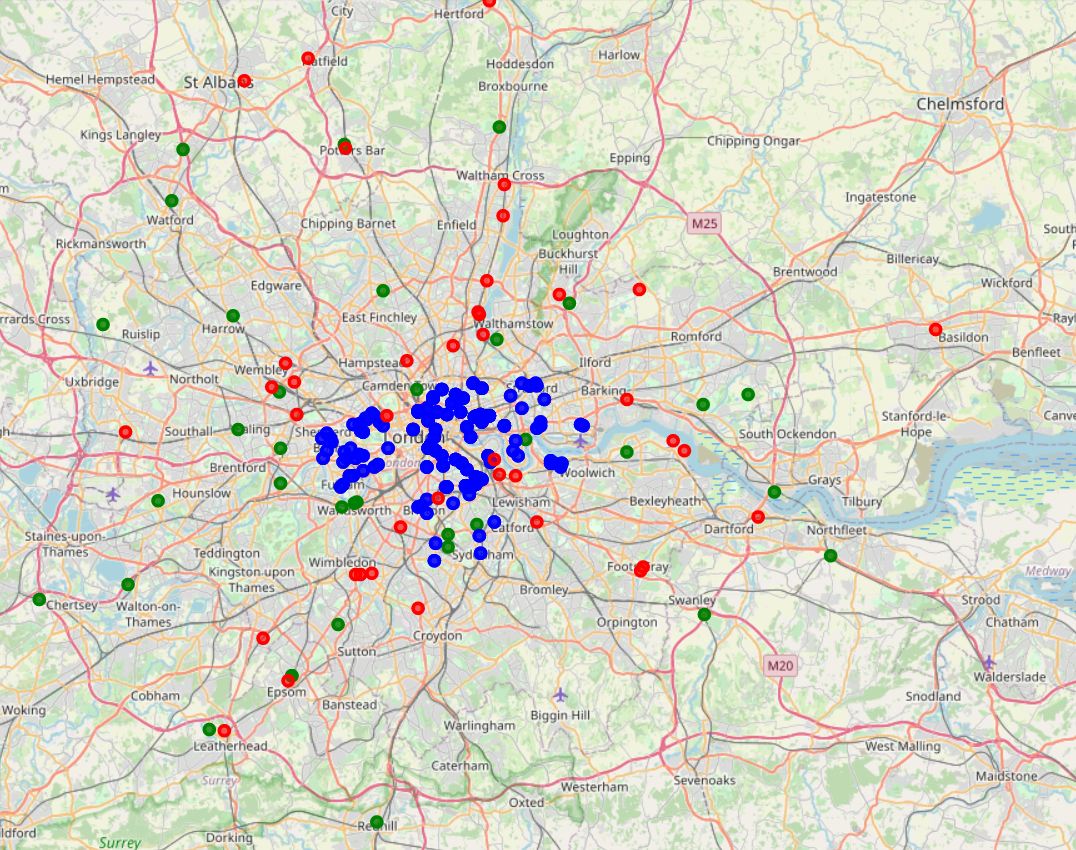
 

Lagos population by Local Authority Summary

<https://africacheck.org/reports/what-a-waste-fact-checking-four-claims-about-nigerias-garbage-problem/> states some difficulty with getting reliable data on Lagos for waste and population.

**London Recycling Facilities, Supermarkets and Factories:**

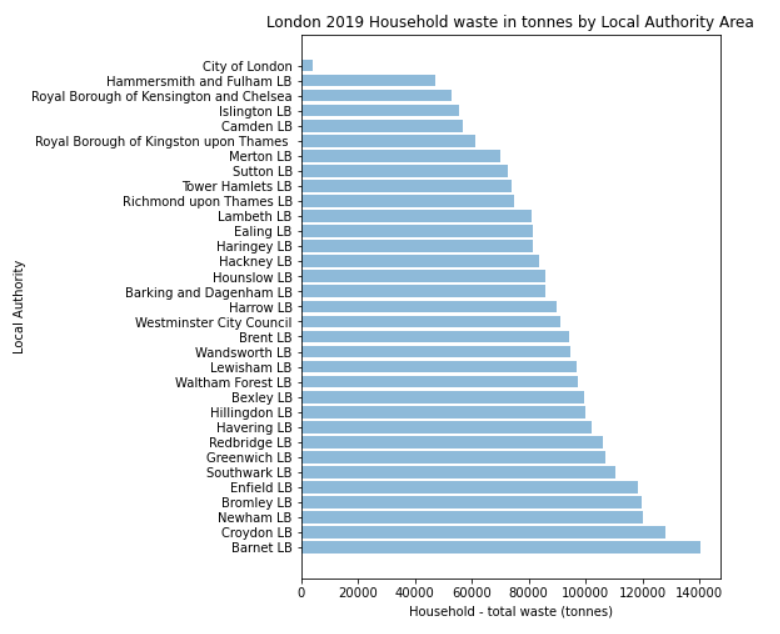
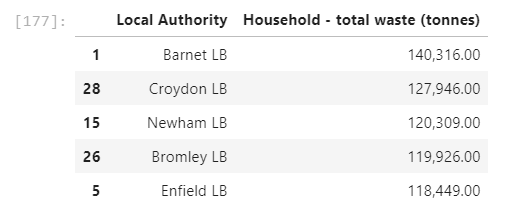
There are 35 recycling facilities (green dots) within 30km radius of Tower of London. There are a total of 170 supermarkets (blue dots) within the same 3km radius of Tower of London, Harringay, Kensington, Canning Town, Dulwich Village. Also, there are 40 industrial estates (red dots) within 30km radius of London (Tower of London). Note that the search for supermarkets was only limited to 3km as any larger radii gave results much larger than the limit permited by the API calls. So, multiple searches were conducted using different central locations at each point. This means some supermarkets falling between two radial points may not be captured.



London Folium map showing recycling facilities (green), supermarkets (blue) and ind. estates (red)

**London Waste and Population by Local Authority:**

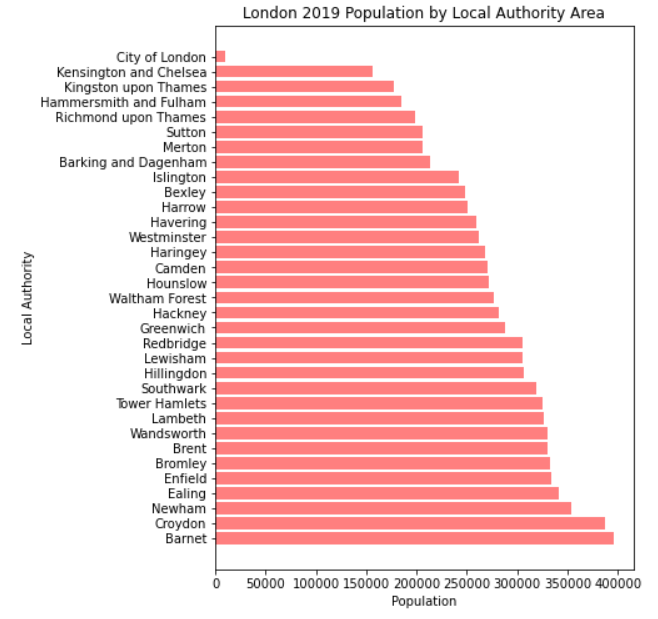
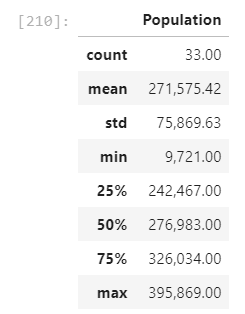
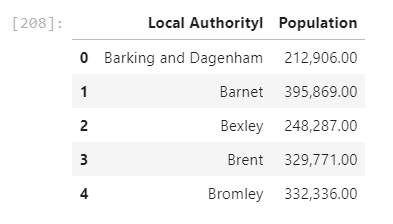
As shown below, there are 33 Local Authorities/Boroughs in London producing a total of 2.88mio tonnes of waste per year (based on 2018/2019). This is an average of 87k tonnes per LA per year. Barnet produced the largest waste of 140ktonnes whilst City of London generated the least of about 3.8ktonnes per year.



London Household Waste by Local Authority, 2019

The results below show London’s 33 LAs with a population of 8.96mio people in 2019, averaging 271k per LA, with Barnet as the largest with 395k people and City of London with only 9.7k people.



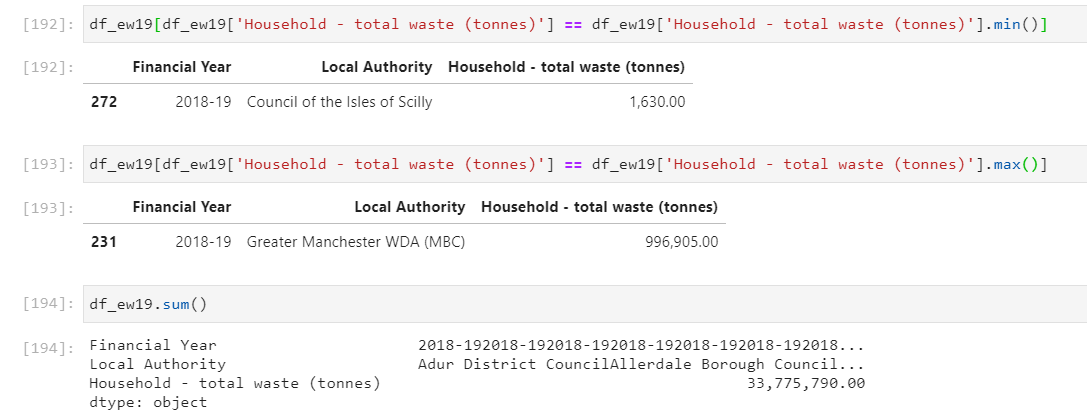
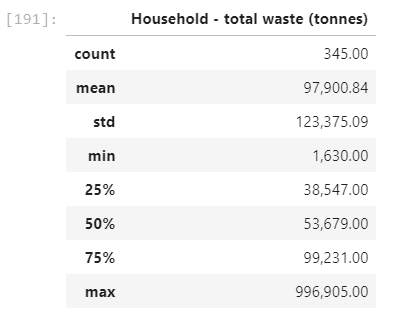
London Population by Local Authority, 2019

Like the data for waste, Barnet and City of London are the largest and smallest respectively in population and waste generation. This may suggest likely correlation between population and waste.

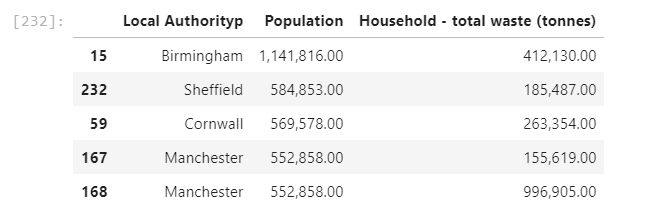
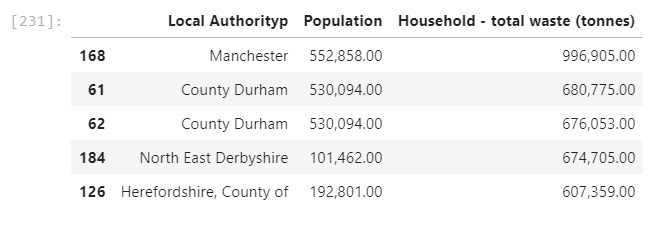
**England Household Waste and Population by Local Authority**

Based on a dataset that contained waste data for all 345 England LAs, all England LAs generated 33.8mio tonnes of waste in 2018/19, making an average of 98k tonnes per LA. Manchester with a population of 553k people generated the largest waste of 997k tonnes, whilst Isles of Scilly generated only 1.6k tonnes. Meanwhile Birmingham with a population of 1.1mio people generated 412k tonnes of waste which is less that half of Manchester’s waste. This suggest perhaps a not so linear relationship at a National (England) level.

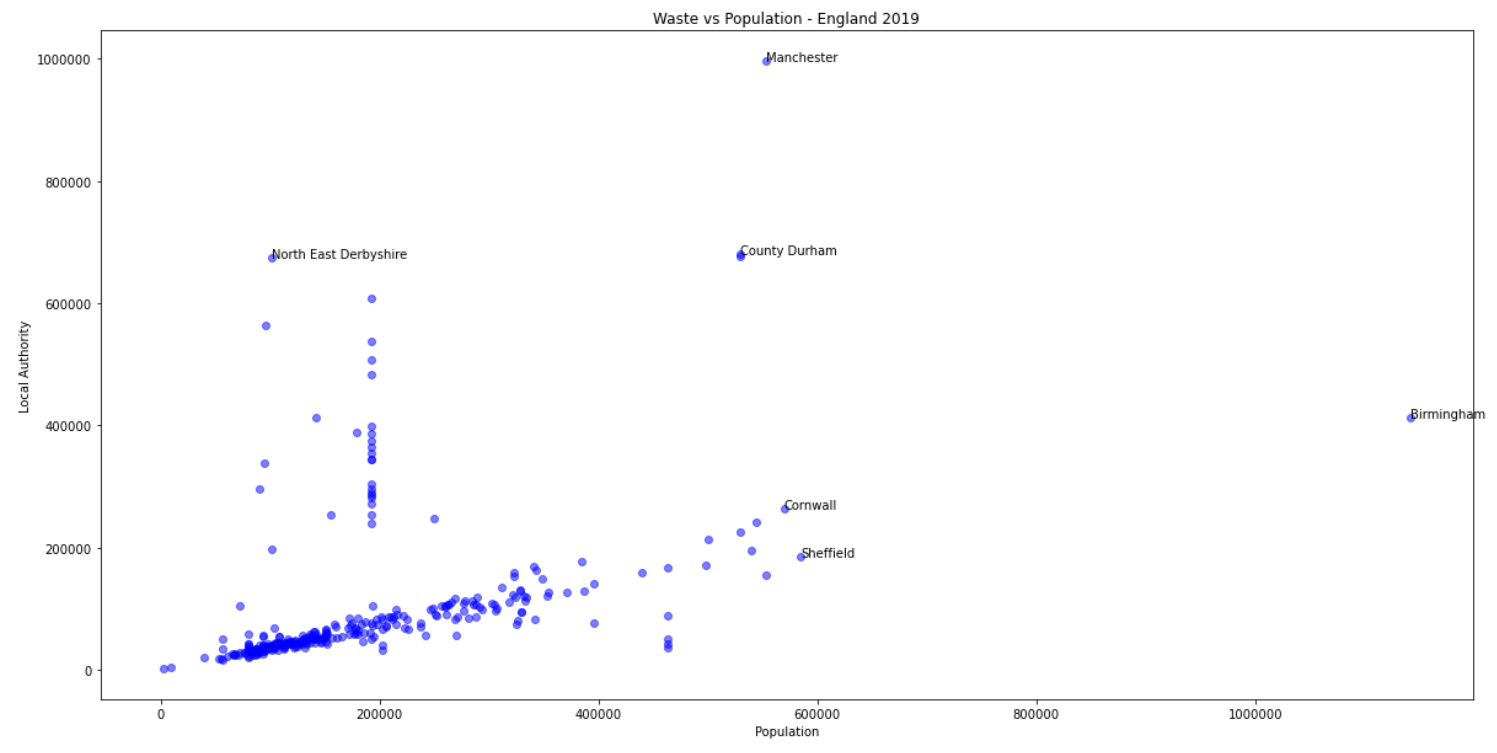
It is also worth noting that the actual England data used for regression and cluster analysis featured only 314 of the 345 Local Authorities for which both waste and population 2019 data for 2019 were available. The merged table also had some errors or repeat Counties. It is a similar output as the summary also features the min, max and very similar average to the full dataset.



England Waste Summary based on original waste data for all 345 England LAs.

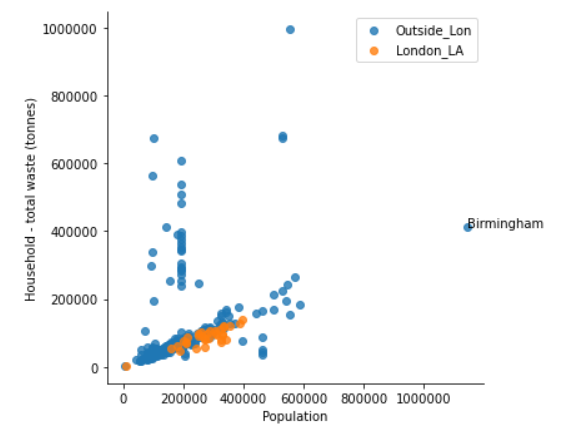


England Population and Waste Summary based on merged table of 314 England LAs.



Scatter Plot of England Waste vs Population highlighting some outlier LAs

The plot below showing London LAs relative to other non-London LAs suggest that London LAs are actually not amongst the largest waste generators in England and perhaps at least in London, a linear relationship likely exists between the population and the amount of waste generated per annum., at least based on 2019 data.



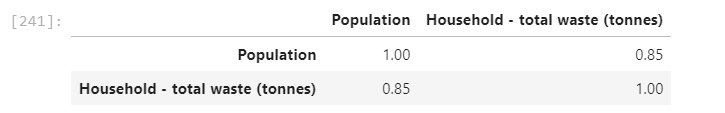
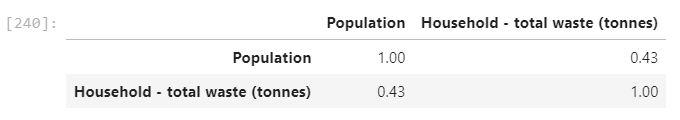
Scatter Plot of England Waste vs Population highlighting London LAs

Good quality data does not exist for wastes in Lagos. And even with API calls and google searches, it is easy to conclude that the 35 recycling facilities within 30km of Tower of London are close to double the 19 recycling facilities within 30km radius of Lagos Isolo, despite the 30mio population of Lagos which more than triples London’s 8.9million people. In a similar way, London’s 170 supermarkets (within a much smaller radius of about 6km) are way more than the 24 found within a wider 30km radius of Lagos. Lagos probably has more open markets which were not captured in the API calls/database. Lagos has slightly more but similar amounts of industrial estates of 47 compared to London’s 40. All these give some perspective that Lagos is under resourced for waste handling and recycling, but London may be more prone to generating higher packaging wastes as there are many more supermarkets whose key model is based on retailing packaged produced, whereas Lagos’ fewer supermarkets and more open markets could make potentially generate comparably fewer packaging waste as those markets are more based on direct farm o market models of retail where packaging needs are lesser, although with consequentially higher food waste. These effects may be cancelling each other out between Lagos and London, but without comparably good quality data for both cities, it is difficult to so and this influenced the decision to leave out these factors from the subsequent analysis and models.

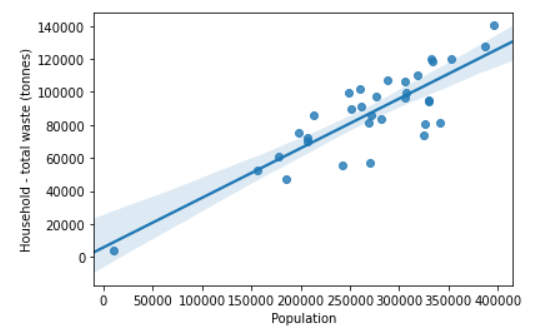
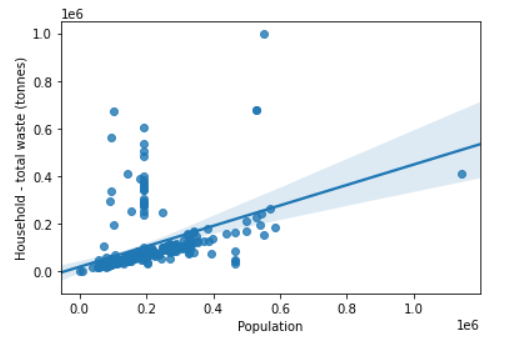
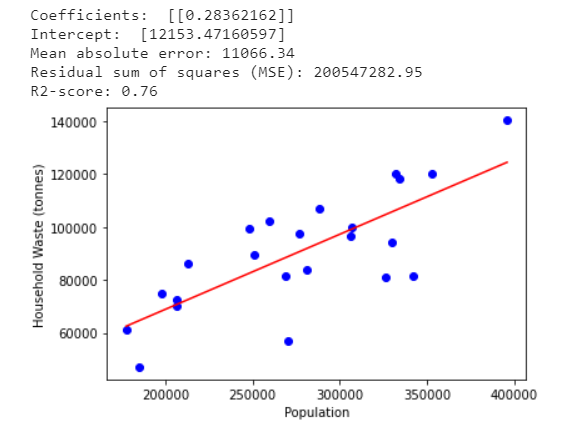
**Regression Analysis:**

*Linear Fit*

The correlation table below shows that there is a reasonable positive linear relationship between population and waste generated in London with a Pearson Correlation Coefficient of 0.85. But this is less so at the England National level with a PC coefficient of only 0.43. The very low p values also confirm a high likelihood that the above assumptions are correct. The R2 suggests a reasonable prediction of waste from population using the model. Other fits were explored to check if better results could ensue.





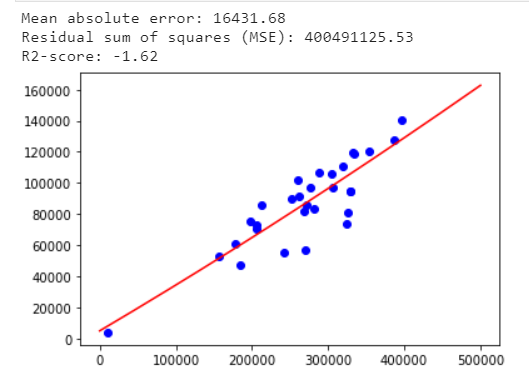
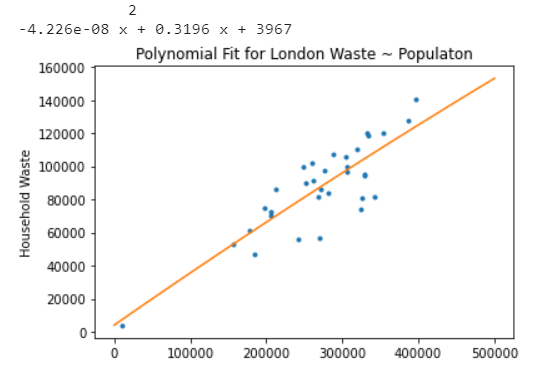
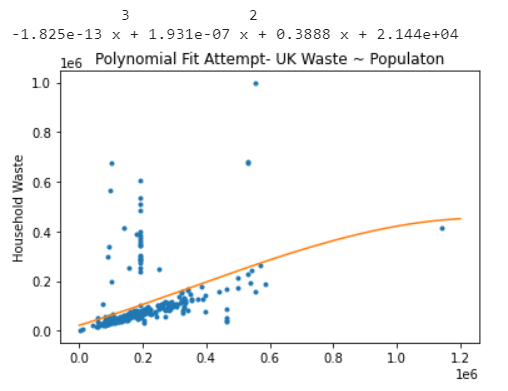
 

Linear Correlations for England and London Population vs Waste

*Polynomial Fit*

The polynomial gave a reasonable fit for London but not so for the England data. In addition, the R2 value which is negative and lower than the Linear fit suggest that a polynomial may be less successful as predictor of waste from population based on the London data. Below are coefficients generated after training and testing the dataset.

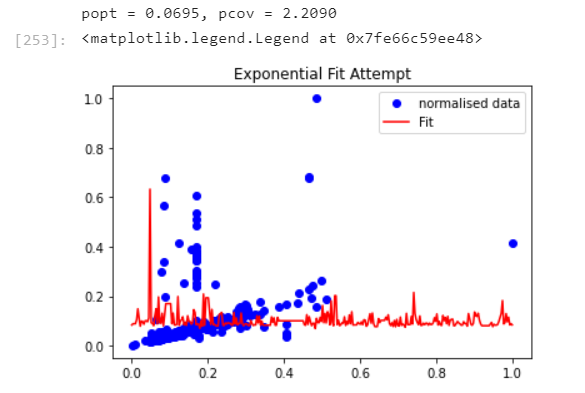
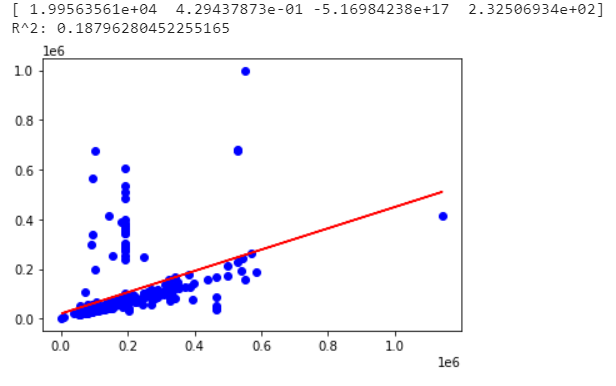




Polynomial Fit for England and London Plots of Population vs Waste

*Exponential Fit*

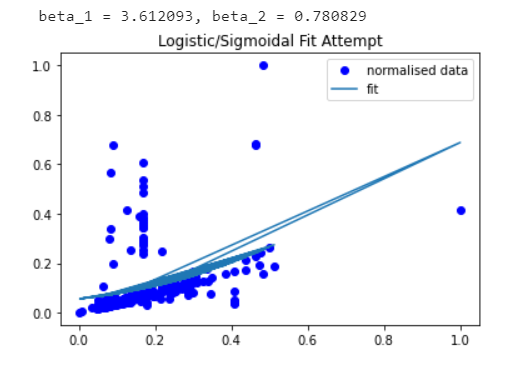
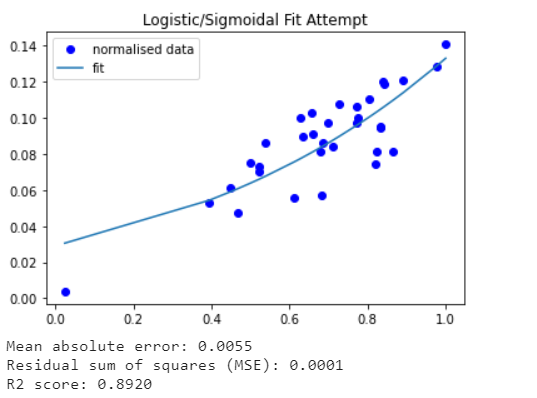
Based on the results below with again poor fit to the England data and low R2 values for London data, exponential model is not a good fit for the data for prediction.

Exponential Fit for England and London Plots of Population vs Waste

*Logistic/Sigmoidal*

Similar to the exponential model, a sigmoidal was attempted for completeness, and as expected did not fit the England data with a surprisingly good fit and R2 value for the London data.

Logistic Fit for England and London Plots of Population vs Waste

Given the reasonably good linear regression fit, the model below can be used to estimate/predict waste for LAs with no or poor data such as Lagos using available population figures.

*Waste/yr. (tonnes) = 0.28\*(City Population) + 12153 error = (+/- 11066)*

Using the above predicts a total waste of 8.9million tonnes of waste for Lagos with a population of 30.5mio people. For context, London’s 8.9mio people generated 2.9mio tonnes of waste in 2019, so perhaps the model may useful for predictions at city levels. It is however important to note that further careful considerations and verifications are necessary, because many other factors that may contribute to improving the model were not used in the analysis. Such factors as lifestyle and affluence, numbers of supermarkets, packaging production factories and recycling facilities within each Local Authority were not used due to restrictions on time and retrieving quality data.

A Finnish waste infrastructure corporation -Woima, in their study of Lagos claim the city generates 13,000 tonnes of waste per day which approximates to 4.7mio tonnes annually. <https://woimacorporation.com/drowning-in-waste-case-lagos-nigeria/>. This is way lower than the projections of the model. But if we considered that the Woima study was based on a population estimate of 20mio people for Lagos, then the model predictions at 5.7mio for a 20mio people city is 1mio tonnes more than Woima’s estimates. Based on World Populations Review’s estimate of Lagos population at 14,368,332, Lagos should generate 4.1mio tonnes of waste. <https://worldpopulationreview.com/world-cities/lagos-population>. Although these numbers do not seem to corroborate the model’s estimates, it is not enough to completely rule out the model given that many of those sources are actually using inaccurate estimates of Lagos population.

**Cluster Analysis:**

The k-means cluster analysis generated the chart and categories of LAs below for the England dataset.

There seems to be 3 main clusters with low waste levels and they are mainly low population:

Low Population LAs (<200k people) with Low Waste (<200k tonnes/year) - (LP LW) – Orange.

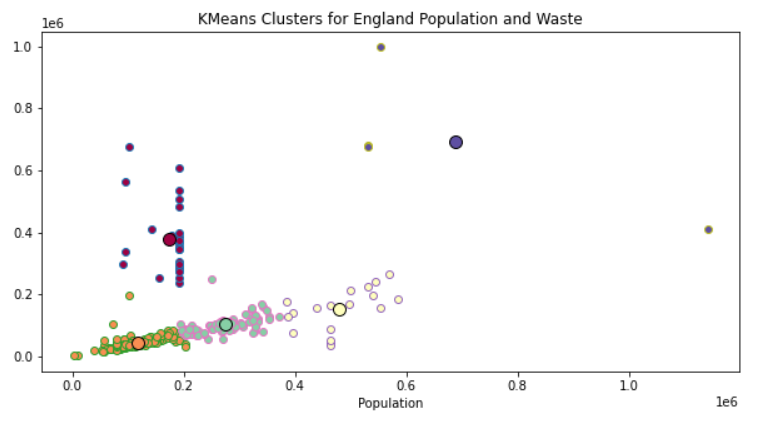
Low Population LAs (<400k people) with Low Waste (<200k tonnes/year) - (LP LW) – Green.

Medium Population LAs (>400k people) with Low Waste (<300k tonnes/year) - (MP - LW) – Yellow.

Also, there are two clusters with medium to high waste levels:

Low Population LAs (<200k people) with Medium Waste (>200k tonnes/year) - (LP MW) – Wine.

High Population LAs (>500k people) with High Waste (>400k tonnes/year) - (HP LW) – Purple.



Clusters for England LAs based on waste and population.

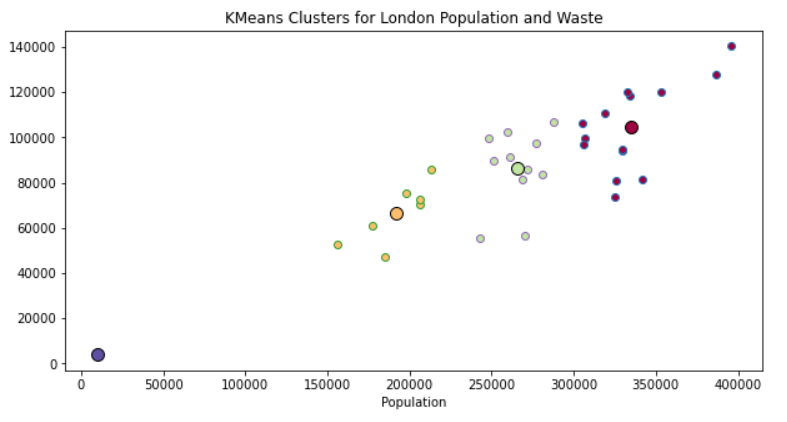
Of the four London Clusters, there is one outlier/lone LA with very low population and waste (Purple). This is the City of London which has extremely low number of households, as it is a mostly financial and banking LA, which explains the low waste levels.

At a national context, all London LAs are Low population and low waste category above, but with a particular focus and scale within London, there seems to be 3 main clusters with medium to high waste levels:

Medium Population (<250k people) with Medium Waste (<100k tonnes/year) - (MP MW) – Orange.

Medium Population LAs (<300k people) with High Waste (<120k tonnes/year) - (MP HW) – Green.

High Population LAs (>300k people) with Medium to High Waste (80k – 140k tonnes/year) – (HP HW) – Wine.



Clusters for London LAs based on waste and population

**CONCLUSIONS AND RECOMMENDATIONS**

**Conclusions:**

Good quality data does not exist for wastes in Lagos, especially not at the Local Authority level with a total of 20 Local Authorities and 30.5mio people. Also, the data for recycling facilities, supermarkets and factories in Lagos, alongside API call limits made it difficult to use such data subsequently in model development, thus the modelling focused on London and England Local Authorities (LAs).

London’s 33 LAs are middle of the pack with respect to population size and waste generation compared to other LAs in England. Barnet was the highest populated with 395k people and generated a waste of 140k tonnes of waste in 2019. For some context, Manchester with 550k people produces approximately 1mio tonnes of waste in 2019. Other high waste producing LAs in England are County Durham with 680k tonnes and North-East Derbyshire with 675k tonnes of waste. Although Birmingham is the most populated LA with over 1.1mio people, it generated only a moderate 412k tonnes of waste in 2019.

In London, there was a linear relationship between the population of Local Authorities and the amount of household waste produced within the Borough. The model below will be useful for such purpose. However, at the National England level it was not possible to find a very good model to fit the relationship between City populations and the amount of waste generated.

*Waste/yr. (tonnes) = 0.28\*(City Population) + 12153 error = (+/- 11066)*

The above estimates that Lagos with a population of 30.5mio will produce about 8.9 tonnes of waste in a year. This may be a useful model to generate reasonable data of waste at Local Authority level for cities where such data may be unavailable or difficult to obtain.

Generally, compared to the national England level, all London LAs are generally low population and low waste producing with a maximum waste at Barnet being just 140k tonnes per year. But bear in mind that population density was not a factor used in this analysis. The LAs to focus upon for improvement in England are those in the cluster of low population but high waste such as County Durham and North East Derbyshire, as their waste production per capita are higher than other LAs.

**Recommendations:**

A next step for this study could be to test out the model on countries with known reliable information on waste and population. This would help check if the model is ideal to use for places with less reliable waste information like Lagos.

A future study should consider incorporating more attributes or parameters (such as numbers of recycling facilities, supermarkets, packaging factories and lifestyle) to perhaps find correlations and models that may be better fits at predicting waste for Local Authorities with better R2 values and smaller errors.

The fuzzy-merge package used to combine tables of waste and population for England seemed to have lost data for some Local Authorities and also generated some duplications that may not have helped with respect to the modelling accuracy.

Other tools like Chloropleth would be useful to visualize the distribution of population and waste across Local Authorities on a heat map.