**LONDON V MANCHESTER FOR AUTO-GARAGES**

**Data**

**This report requires a variety of data obtained through several different sources:**

1. **Population density of the UK per region (*Wikipedia - https://en.wikipedia.org/wiki/List\_of\_English\_districts\_by\_population\_density#:~:text=1%2C000%20-%2010%2C000%20%2F%20km%C2%B2%20%20%20,%20%208%2C636%20%2033%20more%20rows%20)***

**This is the main dataset other values are added to. The full contents were acquired through webscraping, and contains:**

| **Rank** | **District** | **Density(per km²)** | **Type of district** | **Ceremonial county** |
| --- | --- | --- | --- | --- |

**The columns “Rank” and “Type of district” were dropped, and district names converted from objects to strings. To convert the “Density” to an integer, every comma in the data needed to be removed first.**

1. **Geographical areas of each regions (*ONS - https://ons.maps.arcgis.com/home/item.html?id=a79de233ad254a6d9f76298e666abb2b*)**

**This dataset was found on the ONS website, within a zip file in CSV format. To make it accessible, it has been uploaded to a GitHub repository (link available in code). The dataset contains a variety of different measurements for land area, and it was decided it was best to go with the “True” land area, which is the total area minus areas beyond coasts, and any inland water. This is then added to the main dataset, using a code that matched area names, and used to create the cars/hectare column later. There were some areas which had non-matching names, but fortunately none of them are in London or Manchester.**

**After this, the dataframe is split into 2 new dataframes for London and Manchester (based on their “Ceremonial County” entry), and the indexes reset.**

1. **Locations of each region (Geopy.Nominatum)**

**Code was created to go through each location in each dataframe, and add the latitude and longitude coordinates to lists, which were then added to the dataframes. These were then used later to create the maps.**

1. **Garage locations in each area (Foursquare)**

**For London and Manchester, all the auto-garage venue locations were returned in json format, put into dataframes, and then most of the columns dropped leaving only "Name", "Latitude" and "Longitude". These were then kept in separate dataframes to use later.**

1. **Number of cars in each region (*UK Government Asset Publishing Website -*** [*https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/985605/veh0105.ods*](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985605/veh0105.ods)**)**

**The ods file type required pandas\_ods\_reader to load the data into a dataframe. The dataframe was sliced to remove the additional data at the top and the bottom, columns renamed, and index reset. The dataframe datatypes were all “Object”, so the columns “Cars” and “Light Goods Vehicles” were converted to floats, and regional authority names converted to strings. This allowed the values in “Cars and Light Goods Vehicles” to be combined into a new column, which is used to create a new dataframe containing just the region name and the number of cars and vans (this is just referred to as “Cars” from this point onwards). The final change needed was to multiply each value by 1000, in order to convert from 1000’s of Cars to actual car numbers – then this number was converted to an integer.**

**When attempting to combine this data with the London and Manchester dataframes, I found none of the names would match, and I discovered it was because each region name in the dataframe had spaces before the name. To deal with this, every region was renamed using the “lstrip()” function to remove the spaces. Then, it was possible to add all the car numbers to the other dataframes by matching on region names.**

|  | **District** | **Density(per km²)** | **Area(Hectares)** | **Latitude** | **Longitude** | **Cars** | **Cars/Hectare** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Kensington and Chelsea | 12884 | 1212.38 | 51.498480 | -0.199043 | 41558 | 34.3 |
| **1** | Hammersmith and Fulham | 11308 | 1639.76 | 51.492038 | -0.223640 | 46691 | 28.5 |
| **2** | Islington | 16097 | 1485.65 | 51.538429 | -0.099905 | 38340 | 25.8 |
| **3** | Lambeth | 12157 | 2681.01 | 51.501301 | -0.117287 | 68273 | 25.5 |
| **4** | Wandsworth | 9528 | 3426.34 | 51.457027 | -0.193261 | 85314 | 24.9 |

**New columns in the London and Manchester dataframes were created showing cars divided by area of land, which gives a measure of cars per hectare. The index was reset a final time, and values in “Cars/Hectare” rounded to 1 decimal place. The header of the main London and Manchester dataframes look like this:**

|  | **District** | **Density(per km²)** | **Area(Hectares)** | **Latitude** | **Longitude** | **Cars** | **Cars/Hectare** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Stockport | 2315 | 12604.02 | 53.407901 | -2.160243 | 429836 | 34.1 |
| **1** | Manchester | 4735 | 11564.86 | 53.479489 | -2.245115 | 177886 | 15.4 |
| **2** | Bury | 1911 | 9946.01 | 53.592754 | -2.297283 | 114801 | 11.5 |
| **3** | Trafford | 2229 | 10604.47 | 53.418936 | -2.359297 | 121112 | 11.4 |
| **4** | Salford | 2617 | 9719.74 | 53.487746 | -2.289192 | 105636 | 10.9 |