**LONDON V MANCHESTER FOR AUTO-GARAGES**

**Introduction/Business Problem**

I am a data analyst working for a company that would like to open a new auto-garage, in either London or Manchester. The aim is to find the best area within each city to open one, and to see which is the best place overall.

Where to open a new business is a difficult choice, as a business can succeed or fail based on this alone. It is important to know what the best factors to bear in mind when determining this location, and to find associations between variables where there are any.

**The most important thing to consider if opening an auto-garage, is the density of cars in the region (eg. Cars/km2). Population density should not be an important consideration (it only matters how many cars there are), but it will be interesting to see if there are any associations that may cast long term consideration of where the best venue may be. Another important consideration is how sufficiently the area is currently serviced by auto-garages. If an area with a relatively high density of cars is found, the number of existing auto-garages in the vicinity may cancel out any advantage found, or even offset it.**

For the purposes of this exercise, the type of auto-garage to be opened is the most common type that services cars and small goods vehicles, so we will be combining the figures for those and leaving out other vehicle types (motorbikes, heavy goods vehicles etc.).

I will create maps showing the current auto garages in each area (provided by Foursquare), overlayed with a colour sensitive map of each district based on the density of cars, as well as conducting exploratory data analysis to find correlations in data that I hope will give a greater understanding of the factors that result in more cars in each district. This data analysis project will not be measuring cost of land as it would be so variable within each area, but it will be a consideration when selecting locations.

Anyone wanting to open an auto garage in London or Manchester would be interested in this dataset, as well as people wanting to know the density of cars in each region for other reasons such as traffic analysis, or to open a car dealership.

**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 |

**Exploratory Data Analysis/** **Methodology**

**Dataframes were created in order to see if anything is correlated to “Cars/Hectare” in London and Manchester.**

| LONDON | **Cars/Hectare** |
| --- | --- |
| **Density(per km²)** | 0.784428 |
| **Cars** | -0.720960 |
| **Area(Hectares)** | -0.844666 |

| MANCHESTER | **Cars/Hectare** |
| --- | --- |
| **Density(per km²)** | 0.285304 |
| **Cars** | 0.959771 |
| **Area(Hectares)** | -0.196975 |

**A screenshot of a computer

Description automatically generated with medium confidenceChart, scatter chart

Description automatically generatedThe correlation coefficients show some relationships, as well as broad differences between the London and Manchester dataframes. To examine this more closely, 3 sets of regression plots are created in Seaborn:**

**A picture containing text, screenshot

Description automatically generated**

**From looking at these graphs, it is clear that relationships between the data in London and Manchester are significantly different, so it is necessary to keep the data separate to make observations.**

1. **London**

**Graph 1: There is a positive correlation between the number of cars in an area, and its density of population, though it is not a linear relationship.**

**Graph 2: As the overall number of cars in an area increases, the overall density of the cars in an area decreases.**

**A picture containing chart

Description automatically generatedGraph 3: Shows the relationship between Area and Cars/Hectare, and demonstrates that as the size of an area increases, the density of cars in that area decreases. To understand further why this is happening, 2 further graphs are produced for Area vs Population density:**

**Although there is a positive correlation between population density and car density, it is not linear. As the density in an area increases, the gain in the density of the cars which would be expected gets less. This is the reason why the areas with the most cars overall have the least car density, and this is further supported by the “Relationship Between Land Area and Population Density” graph, showing how the most population dense regions have the smallest land area.**

**Conclusion**

**I would surmise that the reason for this, is with the Congestion Charge, cost of parking and particularly good transport links, many people in the densest regions of London do not own a car. There are exceptions such as Kensington and Chelsea which is the most car-dense region in the capital, but those who own a property there probably do not need to worry about money and see owning a car as a symbol of status.**

1. **Manchester**

**Graph 1: There is no relationship between cars per hectare and population density.**

**Graph 2: Although the correlation coefficient between cars and car density is high, from looking at the graph, there is no reasonable certainty that there is much relationship (if Stockport were eliminated from the dataset, it would show a weak negative relationship).**

**Graph 3: Land area against car density shows little relationship, with Stockport being a significant outlier, having a high density of cars compared to its land area.**

**Graph 4: The relationship between land area and population density graph done to further examine the London figures, shows Manchester City has a very dense population for its land area, and there is a weak negative relationship between population density and land area.**

**Conclusion**

**It is interesting to see how the Manchester dataset differs from the London dataset. It seems that density of population has no effect on the density of cars in the area, suggesting that as density increases, the proportion of car use decreases in direct proportion (more so than in London). This may be due to a variety of factors, but is not relevant for further exploration here.**

**Comparing Locations**

**Please see the next page.**

**Chart, histogram

Description automatically generated**

**These bar charts show the car density in each region in London and Manchester. Kensington and Chelsea has the highest car density in London, but Hammersmith and Fulham has a density almost as high, and would have lower land costs.**

**The Manchester chart shows that Stockport has a car density more than double that of the next highest region. This would make Stockport the most desirable place for an auto garage in Manchester.**

**Comparing Stockport to Kensington and Chelsea shows they have approximately the same density of cars.**

**Maps**

**Car density for London and Manchester are combined into 1 dataframe, normalised, split, then assigned back to their respective dataframes. These normalised values are then used to assign a colour to each district based on car density (green for low values, red for high values). These are plotted according to the latitude and longitude coordinates obtained through Geopy.Nominatum, and have pop-up labels with all relevant details about the region. Garages are also visible on the map as black dots, with pop-up labels displaying their names.**

**Map

Description automatically generated with medium confidenceA picture containing chart

Description automatically generated**

**MANCHESTER**

**LONDON**

**Observations for London:**

**There does not seem to be any point choosing a location in the very centre of London, as property prices would be highest, and it does not have a higher density of cars than areas a little further out.**

**The best place in London could be Kensington and Chelsea. Prices in that area would be extremely high, but if an excellent service is provided, it might entice the locals to pay a higher price. There is only 1 garage near the area.**

**Alternatively, the garage could be placed in Hammersmith and Fulham, which is not far from Kensington and Chelsea, would have less land value, and has no auto-garages nearby. In my opinion, the best and safest place would be Hammersmith and Fulham. If it were placed somewhere between, it could attract people from both regions.**

Application, map

Description automatically generated**MAP OF KENSINGTON AND CHELSEA/HAMMERSMITH AND FULHAM**

**Observations for Manchester:**

**The only place there suitable for an auto-garage would be Stockport, as it is the only area in Manchester with a high density of cars. There are 4 auto-garages in the area, but this is not atypical of Manchester and the surrounding areas.**

**MAP OF STOCKPORT**

Map

Description automatically generated

**FINAL CONCLUSIONS**

**While Stockport would seem to be the best place, being the cheapest area and the greatest car density, the 4 auto garages nearby (and others surrounding Manchester) would reduce the amount of business below that which would be available in Kensington and Chelsea, or Hammersmith and Fulham. Therefore, either Kensington and Chelsea or Hammersmith and Fulham would be better. The data also suggests that with the exception of Stockport, people in Manchester use cars less than in London, so in general Manchester would not be considered a good place to set up an auto garage.**

**As discussed earlier, the choice over whether to place your shop in Kensington and Chelsea or Hammersmith and Fulham would depend on the customer base you target, but I would recommend Hammersmith and Fulham first for its high density of cars, its distance from other car repair shops, and while still expensive, would be more achievable than buying land in Kensington and Chelsea.**

***Errors:***

A picture containing text

Description automatically generated**Over the course of conducting this project, I have found limitations, particularly with Foursquare. If I check the locations of garages through google maps, it shows significantly more garages than Foursquare. Also, I found that Geopy.Nominatum retrieved the wrong coordinates for Havering – I checked and found it was nothing to do with my code.**

Text

Description automatically generated

**This is a significant error – it places Havering in central north-west London (near Camden), when it should be well east of London. This error with Geopy.Nominatum was not important for the final analysis – Havering has a low car density and would not have been considered. The missing data from Foursquare however, is very worrying. I investigated if I would get more complete results if I were to change the scope of the project to “Auto-Dealerships”, however that again only contained half of the data there truly was (when double-checking with Google).**

**I have drawn my conclusions based on the data I was asked to use, but venue data in this project would need to be checked with google venue location data before a definite conclusion can be drawn. In most circumstances, I would have taken this to mean that I should abandon the Foursquare API and use Google’s instead, or pick another topic not related to venue data, but we have specifically been instructed to use venue data from Foursquare for this project. It is certainly something I will bear in mind for future projects, and I consider that learning how to use Foursquare will help me when utilising other services.**