DATA 70202- APPLIED DATA SCIENCE

Meeting Housing Need and Building More Homes on Brownfield Lands

Student ID:

10373525

10922581

10691215

10909185

19/05/2022

Word Count: 6992

Contents

1. INTRODUCTION

Brownfield sites are areas that has previously been utilised, nowadays they are abandoned or dilapidated. At first, these sites may be less expensive to purchase, existing structures must often be demolished, and land decontamination costs may be high due to previous industrial use [1]. Brownfield land holds lots of potential space for large amount of housing and reduce building on new spaces, therefore building on brownfield land increases the number of green spaces in country. Brownfield land registers are lands that is appropriate for the residential development and has officially registered for housing. Brownfield land register consists of two parts. The first part is brownfield sites register must done official procedure and will include all brownfield lands that have been determined to be suitable for residential development by a local planning authority. The second part is local authorities give a permission for brownfield land register residential [2]. People will expect comfort and convenience when they start to build a house. Therefore, we should consider the education, leisure centres and transportations, and these are the factors affect the brownfield land register for housing. These weight and importance of factors will be discussed in the next sections.

In recent years, most of the houses can be built on brownfield land, however some of the brownfield land datasets are not in the appropriate form to give information about the development on brownfield land registers. We are looking for the dataset in Greater Manchester. Most datasets found from GeoServer and authority's official website. The main aim for this report is to find enough lands for building more houses. There is digital data about brownfield land and the factors for deciding the availability to build a new house, however dataset is in various data source, so we do not know the available land for more houses in Greater Manchester. In addition to this, dataset is not in the same format such as some of them are csv files and the others are shape files so we cannot compare and decide the availability of the land for new houses. At this stage, this is our main concern to gather all dataset in one location and convert the dataset in same format and be able to compare them.

In this project, brownfield land register data are split into each group members and every member face with some difficulties in their dataset. Some of the dataset has various type of location for instance, east and west rather than latitude and longitude for brownfield land register. The programme ArcGIS pro is used for illustration of the data for brownfield land register. Latitude and longitudes are acknowledged by the programme; therefore, programme detect the location of the places in the map. The map in ArcGIS pro is showed the availability of the land for housing to build. To understand the programme and Fig out the programming language is the second most challenging part for us. The first one most challenging thing is downloading the data for brownfield land register and for each factor. We are focusing on creating automated maps which should be showed the latest version of the available brownfield land for housing. Python is used for coding part for the report which gave us an opportunity to download the latest version of dataset from website, therefore if the dataset is edited as time pass our dataset is also updated. Some of the dataset for the brownfield land and factor part do not publish, only CSV files can be found so the CSV file is downloaded and linked by python code which will be discussed in next sections. The downloaded data are up to date data which gave us an opportunity about automated workflow for brownfield land. To sum up, after relevant datasets are collected, they are converted into same format by using python then all information is import into the programme ArcGIS pro so we can be able to visualise the brownfield land.

2. WORKFLOW

The figure below is the overall workflow of the project, which consists of four steps in total. The first step involves analysing project requirements and finding relevant data, which is the basis of the entire project. Aside from the first step, the goal of the remaining parts is to form the final automated analysis tool collectively.

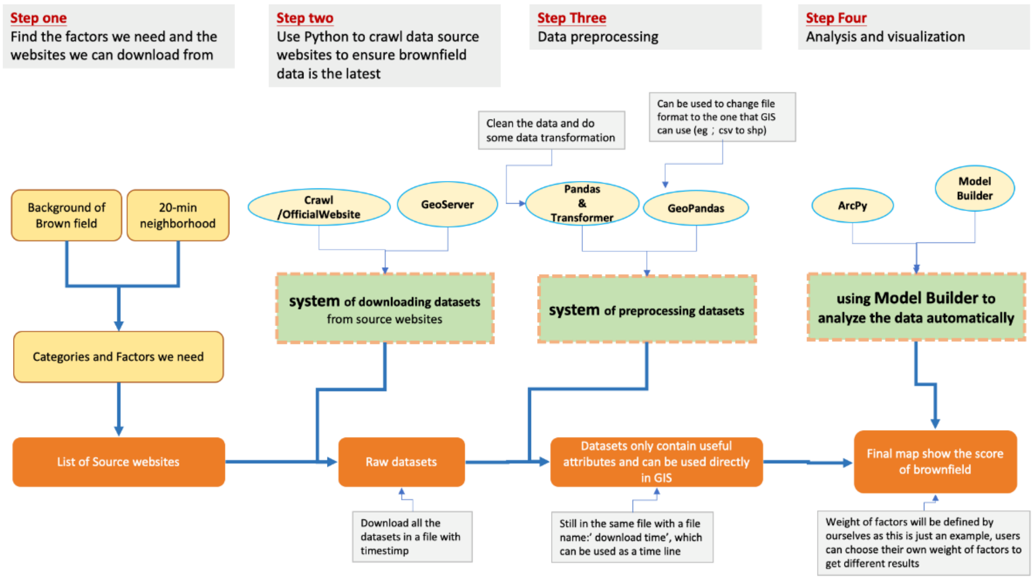

Fig 2.1 Workflow.

The first step is to analyse the project requirements in-depth to determine the required data and select appropriate data sources. It mainly includes brownfield data and some factors that determine whether brownfield is suitable for residential development, such as education data, traffic data, etc. The second step is to use the python program to automatically download the data that needs to be downloaded, store it in a local folder, and use the timestamp to manage the historical data so that users can find the corresponding historical data. This part of the program will use the related technologies of crawlers and the associated details of file path management to prepare for subsequent data pre-processing. Our goal is to automatically download the required dataset from the network each time the program is run. The third step is to pre-process the downloaded data, using python to ensure that the data content and type are what we need. This part includes the following three pre-processing methods such as convert the file format, decompress the compressed file, select useful information and so on. The fourth step will use the built-in Model Builder of ArcGIS to build an automatic spatial data processing and analysis system. Although all the data has become the SHP format after the previous steps, it still needs to be clipped and other processing methods to reduce the amount of calculation in the data analysis. The process of data processing and analysis is created by the Model Builder, which can be exported as an Arc Toolbox box for secondary use and secondary modification.

The above is the overall workflow of the project. The main results will be a python file (steps 2 and 3) and an Arc Toolbox file (step 4). The specific operations and methods of each part will be introduced later.

3. DATA

Brownfield lands datasets are combined from various websites, and they have different attributes and format for the location part which is the most important column in this project. The ArcGIS pro programme pave the way for the location of brownfield land however, we have faced with two main problems. The first one is that the shape files are automatically opened with the ArcGIS programme and did not need any data pre-processing steps. As mentioned in introduction part, some of the datasets are in excel format (csv) so these datasets should be converted into shape file format by using codes in Python. The places for brownfield land register in Greater Manchester are Oldham, Rochdale, Salford, Bolton, Manchester, Bury, Trafford, Wigan, Stockport, and Tameside. There are ten places considered for Greater Manchester, most of the dataset for those places have similar structures. The programme which is called ArcGIS pro can easily recognise the latitude and longitude for the location of brownfield land register, however some of the brownfield land dataset, for instance, Trafford and Rochdale have different type of coordinate system. These diverse types converted into GeoX and GeoY or latitude and longitude format which can be recognised by ArcGIS pro programme. All brownfield land should be on same format so we can compare them with each other. The report is about creating an automated workflow for building house on brownfield land, therefore before converting location we should download the dataset from website by using python code. For each dataset of brownfield land register, there are lots of columns with lots of information however only some of the columns are important for our project. After some research, specific columns are selected for brownfield land dataset such as, ‘Authority,’ ‘SiteNameAddress,’ ‘GeoX,’ ‘GeoY’ and ‘Hectares.’

After finding and downloading the brownfield land dataset, we have considered the most important things for the people before buying and developing a house. To understand the circumstances under which Brownfield is suitable for redevelopment as housing, we mainly refer to Annex 2 in '20-Minute Neighbourhoods': 'Features of a 20-minute neighbourhood –checklist [3]. After discussion and selection, we selected six categories, each of which has some specific factors that affect our choice of housing land. The goal of the 20-minute neighbourhood is to create appealing, fascinating, safe, walkable surroundings in which people of all ages and fitness levels are delighted to move actively for short distances from home to the sites they frequent and the services they require daily. The most crucial factors in our perspective are education, health and security facilities and transportation. Factors also have their subsets, for example, education is divided into nursery, primary, secondary, and high schools. Whole subsets for education need to be considered so we can find the dataset for each subset and convert them into same format, this process is done for each factor. In addition to this example, transportation has several types like train, bus, cycle in Greater Manchester therefore factors should be discussed and decided carefully to decide if the brownfield land register is suitable for housing. Entertainment and leisure facilities, environments and job opportunities are also considered as factor. The reason for this factor is that if the brownfield land is on flooding area most people need to be enduring this situation and build their home by protecting the house from flooding or just do not use that area for housing because it will flood. Local community should take under consideration in all aspect of life and be able to get a comfortable living space.

The weight of factors will be explained in next sections, every factor does not have same importance. We plan to calculate the score of every brownfield of building homes however the weights are decided by AHP method.

4. METHODOLOGY

4.1 Methods of downloading data

Python is a simple language that can be used to automate tasks in ArcGIS products as well as provide good third-party libraries and a wide selection of crawler frameworks. Therefore, we selected Python to write the script for our project.

4.1.1 Download brownfield data by web crawling

The local planning authority updates the brownfield land registers on a regular basis. We opted to crawl that file on each authority's official website to ensure that the data utilised in our final map was current. Crawling works by simulating a computer sending a request to a server, accepting the content of the server's response, parsing it, and extracting the data needed.

In our project, we followed the flowchart explained in Fig 4.1 to crawl 10 authorities' brownfield land registers files in the Greater Manchester. We first need to input a user's header to disguise as a browser. Then we called the 'scrapy' function customized for this project to crawl the file's URL. This function has three parameters, as indicated in Fig 4.2: ‘URL’, ‘select’, and ‘headers. ‘URL’ means the given link of each website. ‘select’ refers to the file's location on the page, which may be accessed via Copy Select in Chrome. To begin, use the Requests [4] library’s *‘get ()’* method to request the supplied website and retrieve the page's data. Then, using the BeautifulSoup [5] library, parse the page source code requested by the Requests library into Soup documents for filtering and data extraction. The BeautifulSoup library has four primary parsers, and we choose html. Parser since it is a built-in standard library for Python that has a reasonable execution speed and excellent document fault tolerance. After getting the URL, we must check whether it is complete; if not, we must add a specific prefix, such as the 'http://www.rochdale.gov.uk' Prefix for Rochdale to make it a valid link; if yes, we need to determine whether it is the file's URL; if yes, we download this file using the *‘urlretrieve ()’*method; if no, we must call the 'scrapy' function again with the new URL, such as Oldham, which has two subpages, requiring three calls to get the exact URL of the file.

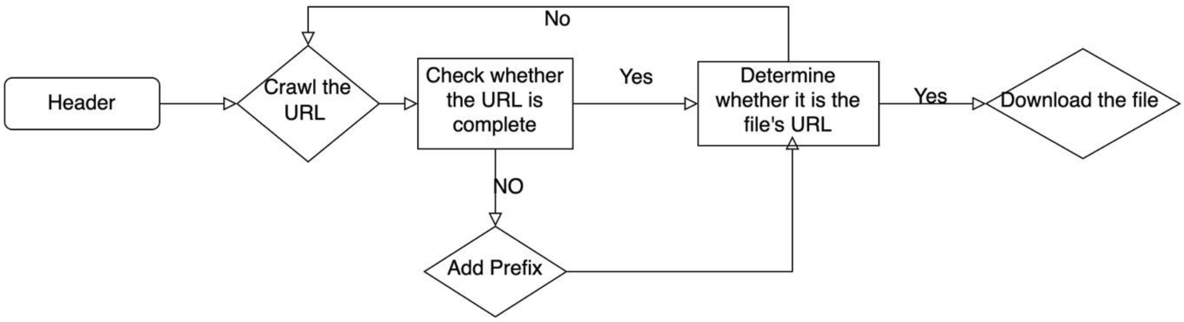

Fig 4.1 The flowchart of web crawling.

```
def scrapy(url, select, headers):
    res = requests.get(url,headers=headers)
    soup=BeautifulSoup(res.text,'html.parser')
    url=soup.select(select)
    file_url=url[0].get('href')

    return file_url
```

Fig 4.2 The code of the ‘scrapy’ function.

4.1.2 Download factor data from GeoServer

GeoServer [6] is a free and open-source application that allows users to share spatial data. It also has user interfaces for inserting, updating, and deleting geographic data. OpenLayers and GeoServer together provide an open-source solution for our mapping needs, allowing us to find the data we need using OpenLayers. The data in factors is updated less regularly and using the latest version of the data is not required. Furthermore, the publisher publishes SHP files directly in GeoServer, which can be easily imported into ArcGIS Pro without data pre-processing, so we download the data via the URLs to the files in GeoServer.

We have selected a total of 9 factors as our scoring criteria for housing land, except for the factor bus-station, the data for the other 8 factors and boundaries of the UK can be found directly in GeoServer with links to the corresponding SHP files, then we used urlretrieve() method to download those files. Because all files downloaded in GeoServer are in zip format, we need to write a function called unzip_file with three parameters to decompress files by unpack_archive() method in Shutil library which can realize some advanced functions for copying, moving, compressing, and decompressing files, as shown in Fig 4.3. For the factor bus-station, we only found the file in CSV format on the official website, so we still used the crawler to get the URL of the file and then downloaded it.

```
# Create a function to unzip .zip files
# 'directory' refers to the directory where the zip file sits. ' extract_dir' is the diretory for unpacked files.
def unzip_file(directory, filename, extract_dir):
    zip_file = directory + filename
    archive_format = 'zip'
    shutil.unpack_archive(zip_file, extract_dir, archive_format)
```

Fig 4.3 The code of the 'unzip_file' function.

4.2 Data pre-processing

Except for those with SHP formatted factors' datasets, all brownfield land registers datasets and the bus station dataset had to be pre-processed by first selecting the required features, then unifying the coordinate reference system in all files to the World Geodetic System 1984 (WGS84), and finally converting the files to SHP format for later use in ArcGIS pro. Fig 4.4 shows the flowchart for this section.

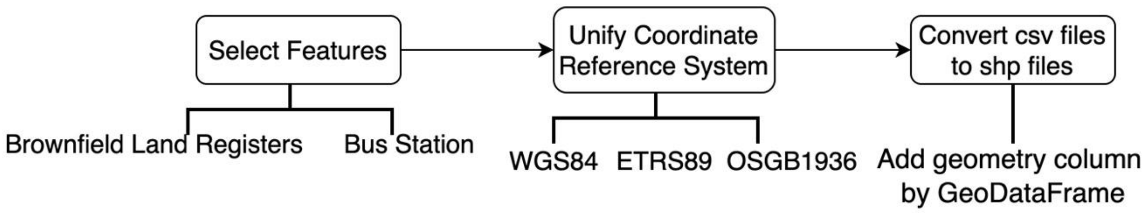

Fig 4.4 The flowchart of data pre-processing.

4.2.1 Features

We simply need to know the address, geolocation, and area of each brownfield land; therefore, we chose four features: 'SiteNameAddress,' 'GeoX,' 'GeoY,' and 'Hectares'. However, some specific instances must be handled because some authorities' coordinates, such as Rochdale's, are not designated GeoX and GeoY. To distinguish the authority in which it is located, a new feature called Authority was created. Users can check the raw data we downloaded if they wish to learn more about a land. Because we just needed the name and coordinates of each bus station to locate it on our map, we only used three features: 'CommonName,' 'Longitude,' and 'Latitude.'

4.2.2 Convert coordinate reference system

- **Reasons for converting coordinate reference system**

WGS84, ETRS89, and OSGB1936 [7] are three coordinate reference systems used in all our datasets. World Geodetic System 1984 (WGS84) is a three-dimensional Cartesian coordinate system with an associated ellipsoid designed for placing anyplace on Earth. WGS84 positions can be defined as either XYZ Cartesian coordinates or latitude, longitude, and ellipsoid height coordinates.

The European Terrestrial Reference System 1989 (ETRS89) is a European adoption of a specific WGS84 epoch to eliminate the effects of tectonic motion. Because of the extremely regular, predictable, and accurate tectonic plate motion, locally fixed GPS datums are very easily and precisely connected back to WGS84, meaning that coordinate translation between these two systems is unnecessary.

Instead of theodolite triangulation, the British National Grid (OSGB1936) is determined using GPS plus a transformation, allowing the use of easting and northing coordinates in UK. Because OSGB1936 and WGS84 are constructed using different methods, we must convert the former to the latter.

There are three reasons why we should use WGS84 as our coordinate reference system. To begin, if it is not uniform, we will project the wrong map when importing all data in bulk. Second, the base map we have chosen is in WGS84 coordinates. Lastly, we wish to use this project's codes and methodology to evaluate housing land alternatives all over the world; ETRS89 is more applicable to Europe, OSGB1936 is only applicable to the United Kingdom, and only WGS84 can be used to locate everywhere on the planet.

- **Converting coordinate reference system by Transformer module**

The code in Fig 4.5 demonstrates how to transfer values from the OSGB1936 system to the WGS84 system. Three steps make up the transformation process [8]:

- Import the Transformer package which allows translation between any pair of definable coordinate systems.
- Make a Transformer from from_crs() function with three parameters. The projection of input data is the first. The output data projection is the second. When the value of always_xy is true, the procedure will accept and return coordinates in the standard GIS (Geographic Information System) order, which is longitude, latitude for geographic CRS and easting, northing for most projected CRS.
- Call the transform () method of a transformer we created to convert the values of easting and northing under OSGB1936 system to the corresponding values of longitude and latitude under WGS84 system.

```
# 3. Convert XY to GeoX,GeoY (Only Trafford and Rochdale )
from pyproj import Transformer
transformer = Transformer.from_crs("epsg:27700", "EPSG:4326", always_xy=True)
transformed=transformer.transform(X,Y)
```

Fig 4.5 The code of converting coordinates from OSGB1936 to WGS84.

4.2.3 Convert CSV files to SHP files

Shapefile is a vector graphics format that saves the location of geometry and related attributes, which can be analysed in GIS, such as drawing buffers, cropping, etc. That is why we need to convert CSV files to SHP files. The code in Fig 4.6 shows three steps to realize the conversion.

- Import the GeoDataFrame [9] and Point packages firstly. A GeoDataFrame object is a pandas.DataFrame that has a column with geometry.
- Create tuples of geometry by zipping longitude and latitude columns in CSV files.
- Call GeoDataFrame() function to convert DataFrame objects to GeoDataFrame objects with three keyword arguments. The first parameter is the DataFrame object to be converted, the second parameter is the Coordinate Reference System of the geometry, and the third parameter is the geometry column on the GeoDataFrame object.
- Save all GeoDataFrame objects in shapefile format.

```
# 2.2 convert Bus station.csv to shp file
geometry_bus = [Point(xy) for xy in zip(dl_bus.Longitude, dl_bus.Latitude)]
crs = {'init': 'epsg:4326'}
gdf_bus = GeoDataFrame(dl_bus, crs = crs, geometry=geometry_bus)
output_filename_bus = folder_shp + 'bus_station' + ".shp"
gdf_bus.to_file(filename= output_filename_bus, driver='ESRI Shapefile')
```

Fig 4.6 The code of converting csv files to shp files.

4.3 Methods of importing data into ArcGIS pro

4.3.1 Reasons for using ArcGIS pro

We used ArcGIS Pro to create our housing land map for distinct reasons. First, our Python code is 3.x, and ArcGIS Pro's python environment is version 3.x which can successfully run our code; however, ArcMap's python environment is version 2.7, and it must be reinstalled as version 3.x. Second, ArcPy is already installed in ArcGIS pro, but we must first generate a Python version that meets ArcPy's requirements before installing it in ArcGIS online notebook. Third, we can design the analytical model for selecting house land utilizing Model Builder in ArcGIS pro. Finally, having a local server allows readers to edit our code and the Model Builder more easily than using an online server.

4.3.2 Importing data into ArcGIS pro by ArcPy

Before running this part code, we recommend users to open ArcGIS pro application and set up a basic project that contains a single map first. And the code consists of five steps [10], as shown in Fig 4.7:

- Import the ArcPy package which provides a productive way to perform geographic data analysis, data conversion, data management, and map automation with Python. Then check for the current working directory of the project.
- The variable 'arpx' refers to the project currently loaded into ArcGIS pro with the parameter 'CURRENT'.
- arcmap=arpx.listMaps()[n] – Here the 'arpx' variable is used to list the Map which will later show the points and geolocations of the Greater Manchester plot areas for the project. 'n' refers to the index of the Map being used. The Map data is stored in 'arcmap' variable.
- After locating all converted SHP files in the directory, a SHP file will be created with the given directory along with the filename in the present directory.
- With the help of addDataFromPath() method, all SHP files' geolocation references are imported to the current map.

```
#Importing shp files to Map from a certain location
import arcpy

directory=os.getcwd() + '\\'+ folder_shp # in windows

#the path to the project file(.aprx) can replace 'CURRENT'
aprx = arcpy.mp.ArcGISProject("CURRENT")
arcmap = aprx.listMaps()[0]

for file in os.listdir(directory):
    if file.endswith(".shp"):
        filename = os.fsdecode(file)
        shp_file = directory + filename
        arcmap.addDataFromPath(shp_file)
```

Fig 4.7 The code of importing all data into ArcGIS pro.

4.4 Create analysis model by Model Builder

After the previous work, all the data were imported into the ArcGIS interface. This part will use the Model Builder to complete the final analysis of all the data. The specific steps include pre-processing the data in SHP format and using the ‘Raster Calculator’ to calculate the final score of the brownfield, that is, whether it is suitable for housing land. The following will first discuss why the Model Builder is used and why the Analytic Hierarchy Process (AHP) is used to determine the weight, and then the specific analysis process and important analysis tools of the Model Builder will be introduced.

4.4.1 Reasons for using Model Builder

Model Builder is a tool provided by ArcGIS to design, create, and edit spatial analysis models, which can help users visualize the analysis process. The design and use of Model Builder are based on flow charts, so it is very intuitively. This is a huge time saver for both the designer and users. Although the Model Builder can be converted to python files, the Model Builder workflow is still used as the result in this project for the following reasons.

It can improve the convenience of use. For urban planners and most people, it is much easier to run Model Builder than python. And the parameters can be easily modified on the corresponding tool in the Model Builder. Some users also need to modify the analysis ideas and processes to conform to the actual situation of their own city. The Model Builder can do this very conveniently by dragging and dropping.

Secondly, the combination of Model Builder and the previous Python workflow can increase the fault tolerance rate of the workflow. After running the previous code, the user can check whether all the data has been downloaded, or whether all the data has been properly pre-processed. If anything goes wrong, just download the missing files and add the data to the ArcGIS interface. That is, choosing Model Builder can provide users with an opportunity to check the workflow, correct errors, and ensure the smooth operation of the entire workflow.

In conclusion, considering the convenience of future users and the fault tolerance of the entire workflow, we choose to use Model Builder as a tool and encapsulate it as Toolbox for users.

4.4.2 Reasons for using Analytic Hierarchy Process（AHP）

Since the analysis of brownfield needs to integrate the combined effects of multiple factors, a method is needed to determine the relative importance of each factor in the overall evaluation. The degree of importance is called the weight, and the larger the weight value, the more important the factor is in housing selection.

There are two main ways to determine the weights [11]. The first is the subjective empowerment method. Experts provide experience and conduct a comprehensive evaluation of indicators, AHP belongs to this category. The second is the objective weighting method, which studies the correlation between indicators through the impact relationship between historical data or indicators and evaluation results. The second method is the quantitative method, which does not require the decision-maker's business experience, but the calculation is more complicated. Among the two types of plans, we did not choose the second type of method, considering that there is no historical data for reference or transparent criteria for selecting housing land, and the project's focus is on the rationality of the workflow. We choose the most popular AHP analytic hierarchy process in the subjective weighting method.

4.4.3 Pre-processing the SHP data

There are some further data pre-processing need to be done in Model Builder.

- **Merge**

The data for the brownfield comes from ten local authorities in Greater Manchester, with a total of ten documents. But in this project, these brownfield data will be analyzed uniformly, so first, use the 'Merge' tool to merge all brownfield data points into one SHP file named brownfield_Merge.

- **Clip**

Extract the boundary data of Greater Manchester and then use this to clip some SHP (Fig 4.8). This operation is mainly for the clarity of the layout because the project does not need to show the data outside the Greater Manchester area. Of course, this also has an additional benefit: avoid spending too much time calculating data that the project does not need. For example, the Flood Risk layer has a large amount of data, including the entire UK. It will be a time-consuming task to create a buffer for the whole of UK data and convert the buffer to raster.

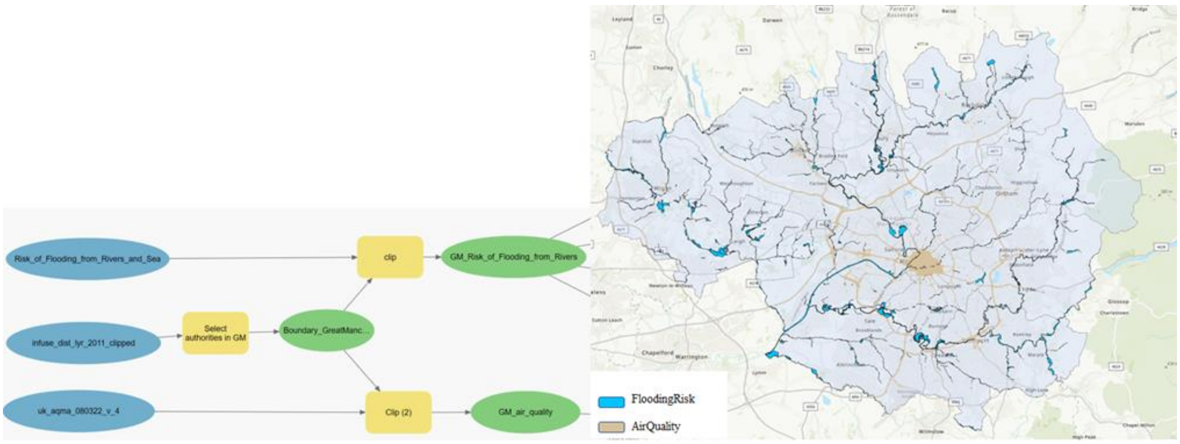

Fig 4.8 Clip the Flooding and Air Quality data, and the result

- **Extract**

Since a single layer also contains multiple features, such as the education and flood layers, feature extraction is required for subsequent analysis. For example, the education layer contains many elements, such as middle schools, universities and primary schools, nurseries, etc. Finally, using SQL statements, the two categories required by the project are extracted, namely, ’secondary school + high school 'and ‘primary + nursery’ (Fig 4.9).

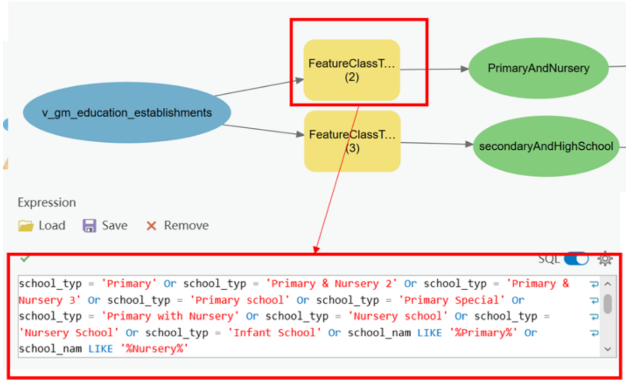

Fig 4.9 Use SQL to extract 'primary + nursery' from education data

4.4.4 Analysis and calculation

In this part, first, create a buffer for some factors, which is the influence area of the factor. Then convert the buffer into raster format and assign the raster value according to the requirements. At last, use the raster calculation formula to calculate the score for Greater Manchester and extract the scores to the brownfield_merge layer.

- **Create Buffer**

Buffer is the most used spatial analysis tool in ArcGIS, which can calculate the buffer range of a given distance based on the original factor. The range of the buffer created is the service range for each factor. However, the problem is that if distinct factors have different buffer ranges, it is difficult to compare the importance when using the AHP method. Therefore, as Fig 4.10 shown, we unified the buffer to 800 meters (due to many bus stops, taking 800 would fill the Greater Manchester area, so take the value 400 meters). Therefore, when using AHP to calculate the weight, it is only necessary to compare which factor is more important within 800 meters.

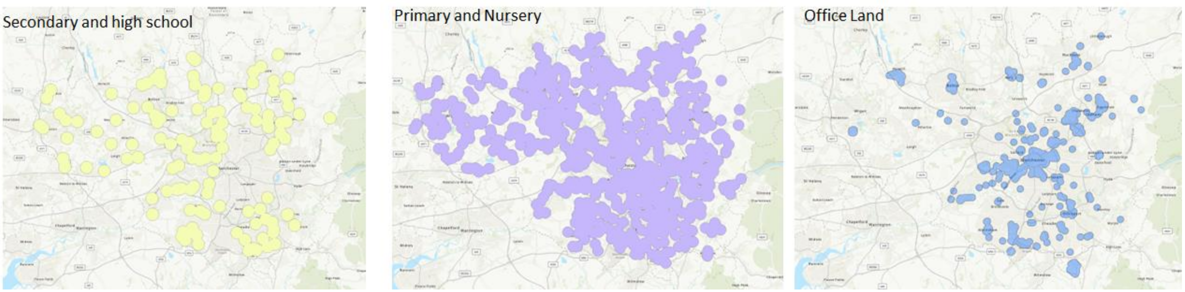

Fig 4.10 Buffers of some factors

- **Convert to Raster**

In ArcGIS, raster data can be used for overlay calculations, but SHP data cannot. So, the buffer layer (belonging to SHP format) created in the previous step needs to be converted to raster data. Values must be assigned to raster cells having varied attributes in the transformed raster， such as BusStation, have a value of 1 for the buffer range, indicating that the bus station can be reached within a given distance; and a value of 0 indicates that there is no nearby bus station. The raster values covered by risk areas in the HighRiskOfFlooding layer have a value of 0 because areas with high flood risk are not suitable for residential use. (Fig 4.11)

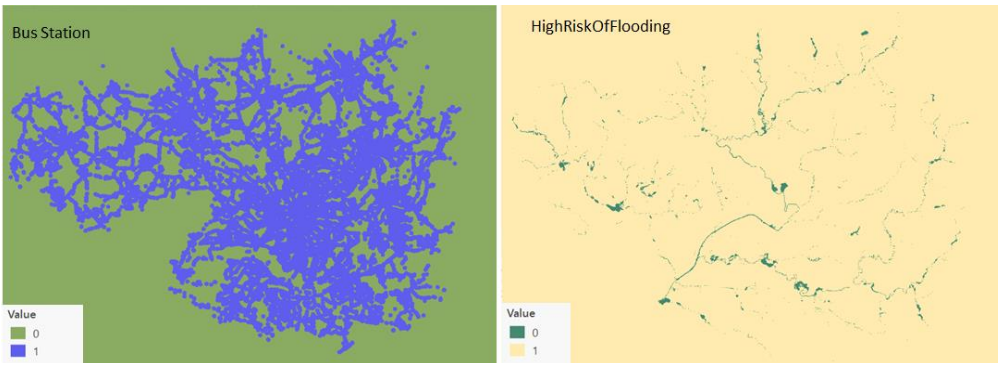

Fig 4.11 Reclass result of BusStation and HighRiskOfFlooding

- **Confirm the Weight**

Before the calculation, the AHP needs to be used to determine the weight of 9 factors as its coefficient. (Flooding risk does not participate in the weight calculation, because we think that the high-risk area is entirely unsuitable for building houses, so the brownfield score of the high-risk area will be directly 0). Select the online AHP calculation tool, the link can be found in Appendix 1. First, the group members discuss and fill in factors comparison form, as shown in Fig 4.12. Choose different numbers to represent the importance of the two factors in the same row. '1' means equal significance, and '9' means the first factor is much more important than the second factor.

With respect to *AHP priorities*, which criterion is more important, and how much more on a scale 1 to 9?

	A - wrt AHP priorities - or B?		Equal	How much more?
1	◉ busStation	○ officeLand	○ 1	○2 ○3 ◉4 ○5 ○6 ○
2	◉ busStation	○ primaryAndNursery	○ 1	◉2 ○3 ○4 ○5 ○6 ○
3	◉ busStation	○ SecondaryAndHighSchool	○ 1	○2 ○3 ○4 ◉5 ○6 ○
4	◉ busStation	○ ChildrenCentres	◉ 1	○2 ○3 ○4 ○5 ○6 ○
5	◉ busStation	○ youthCentres	○ 1	○2 ◉3 ○4 ○5 ○6 ○
6	◉ busStation	○ flooding	○ 1	○2 ○3 ○4 ○5 ◉6 ○

Factors	Weight
ChildrenCenter	0.13
YouthCenter	0.065
PrimaryAnd Nursery	0.17
SecondryAndHighSchool	0.13
Officeland	0.19
CycleHub	0.026
LeisureCenter	0.012
AirPolution	0.027
BusStop	0.25
Total	1
Consistency Ratio CR=0.074<0.1	

Fig 4.12 AHP factors comparison form and weight results

After filling out the form, we can get the weight analysis results (Fig 4.12). From the results, the importance of traffic is the highest, and the importance of LeisureCenter is the lowest. The Consistency Ratio is 0.06, which is less than 0.1, indicating that it has passed the consistency test. This means that there is no apparent paradox in the one-to-one importance comparison step, so the results represent our thinking. Of course, this is only the result of our perspective, and users can use the AHP tool to calculate the weight results from their perspective.

- **Calculate**

The final calculation is done through the tool 'Raste Calculator', and the calculation formula is as follows (SQL statement):

```
("%ChildrenCenter%"0.13+"%YouthCenter%"*0.065+"%PrimaryAndNursery%"*0.17+"%OfficeLand%"*0.19+"%
CycleHub%"*0.026+"%leisureCenter%"*0.012+"%SecondaryAndHighSchool%"*0.13+"%AirQuality%"*0.027+"
%busStop%"*0.25)"%highRiskOfFlood%"
```

In this formula, the product of factors and the corresponding weight are calculated, and sum. Finally, multiplied with highRiskOfFlooding, so that the areas with high flood risk have a score of zero (because in the raster layer of highRiskOfFlooding, the raster value of high-risk areas is 0, and the remaining areas are 1)

- **Final model builder**

The diagram below shows the entire process of Model Builder (see Appendix 2 for high resolution images). The Model Builder has been named '*Brownfield_analysis*' and is stored in the *ADS_Brownfield.tbx* toolbox.

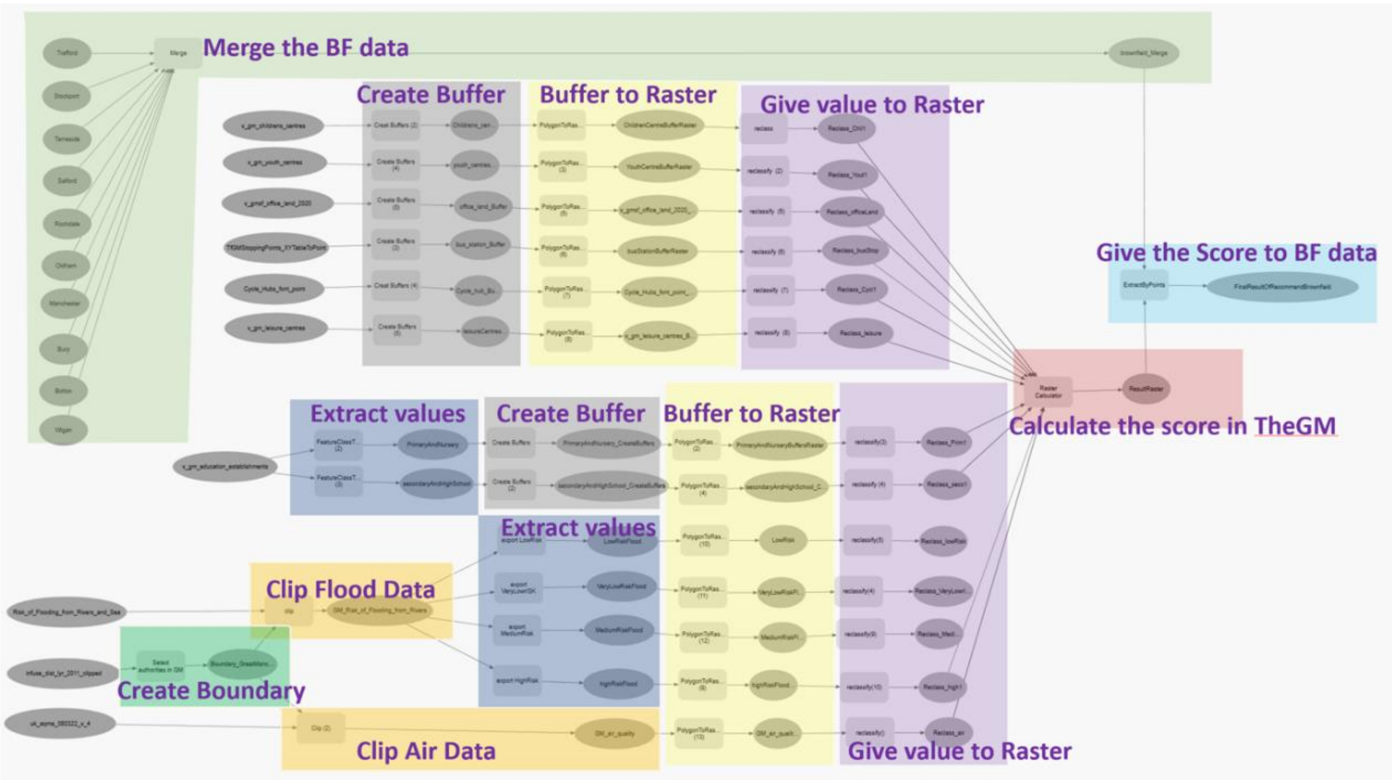

Fig 4.13 Structure of Model Builder

5. RESULTS

5.1 Automatic workflow

Our Python code can be run interactively in ArcGIS Pro using ArcGIS notebooks, which are based on the Jupyter Notebook architecture and provide an experience where code, visualizations, and narrative text can all be executed and saved in one document. The Appendix 1.3 contains the specific process and additional techniques for running code.

Fig 5.1 show how we categorize all files we downloaded and converted into three folders after downloading and pre-processing data part.

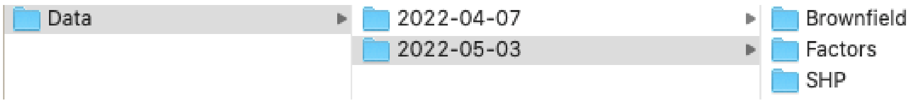

Fig 5.1 The directory we created to store all datasets.

Fig 5.2 depicts the pre-processing process of Trafford's brownfield land registers data. (a) shows the key attributes we chose from the raw data. (b) displays the coordinate values after conversion from OSGB1936 to WGS84. (c) is the conversion of the CSV file to a SHP file with an additional geometry attribute, which can locate sites in ArcGIS pro.

	SiteNameAddress	GeoX	GeoY	Hectares	Authority
0	Stretford Memorial Hospital, 226 Seymour Grove...	395151	381876	0.75	Trafford
1	Former Mosedales Brickworks, 4 Ends Lane, Flixton	393733	374096	0.70	Trafford
2	Higher Road Depot and Adjoining site, Urmston	394716	376968	0.58	Trafford
3	Wharf Road, Altrincham	389103	376754	0.30	Trafford
4	Oakfield Road/Balmoral Road, Altrincham	387828	377186	0.85	Trafford

(a) Raw data of Trafford's brownfield land registers.

	SiteNameAddress	GeoX	GeoY	Hectares	Authority
0	Stretford Memorial Hospital, 226 Seymour Grove...	-2.274386	53.452801	0.75	Trafford
1	Former Mosedales Brickworks, 4 Ends Lane, Flixton	-2.391420	53.439729	0.70	Trafford
2	Higher Road Depot and Adjoining site, Urmston	-2.348259	53.448698	0.58	Trafford
3	Wharf Road, Altrincham	-2.351065	53.398237	0.30	Trafford
4	Oakfield Road/Balmoral Road, Altrincham	-2.344476	53.386796	0.85	Trafford

(b) Data of Trafford's brownfield land registers after converting coordinate reference sysytem.

	SiteNameAddress	GeoX	GeoY	Hectares	Authority	geometry
0	Stretford Memorial Hospital, 226 Seymour Grove...	-2.274386	53.452801	0.75	Trafford	POINT (-2.27439 53.45280)
1	Former Mosedales Brickworks, 4 Ends Lane, Flixton	-2.391420	53.439729	0.70	Trafford	POINT (-2.39142 53.43973)
2	Higher Road Depot and Adjoining site, Urmston	-2.348259	53.448698	0.58	Trafford	POINT (-2.34826 53.44870)
3	Wharf Road, Altrincham	-2.351065	53.398237	0.30	Trafford	POINT (-2.35107 53.39824)
4	Oakfield Road/Balmoral Road, Altrincham	-2.344476	53.386796	0.85	Trafford	POINT (-2.34448 53.38680)

(c) Data of Trafford's brownfield land registers after in SHP format.

Fig 5.2 The results of pre-processing process for Trafford's brownfield land registers data.

We imported all the preprocessed datasets into ArcGIS Pro to create the initial map, as shown in Fig 5.3. Distinct colors represent different elements, as we can see. The grey areas in the diagram are designated Air Quality Management Areas, which are areas where national air pollution regulations have been exceeded. Floods from rivers or oceans can cause flooding in blue areas. All Greater Manchester's bus stations are represented as light orange dots. Other elements can be viewed in the illustration to the right.

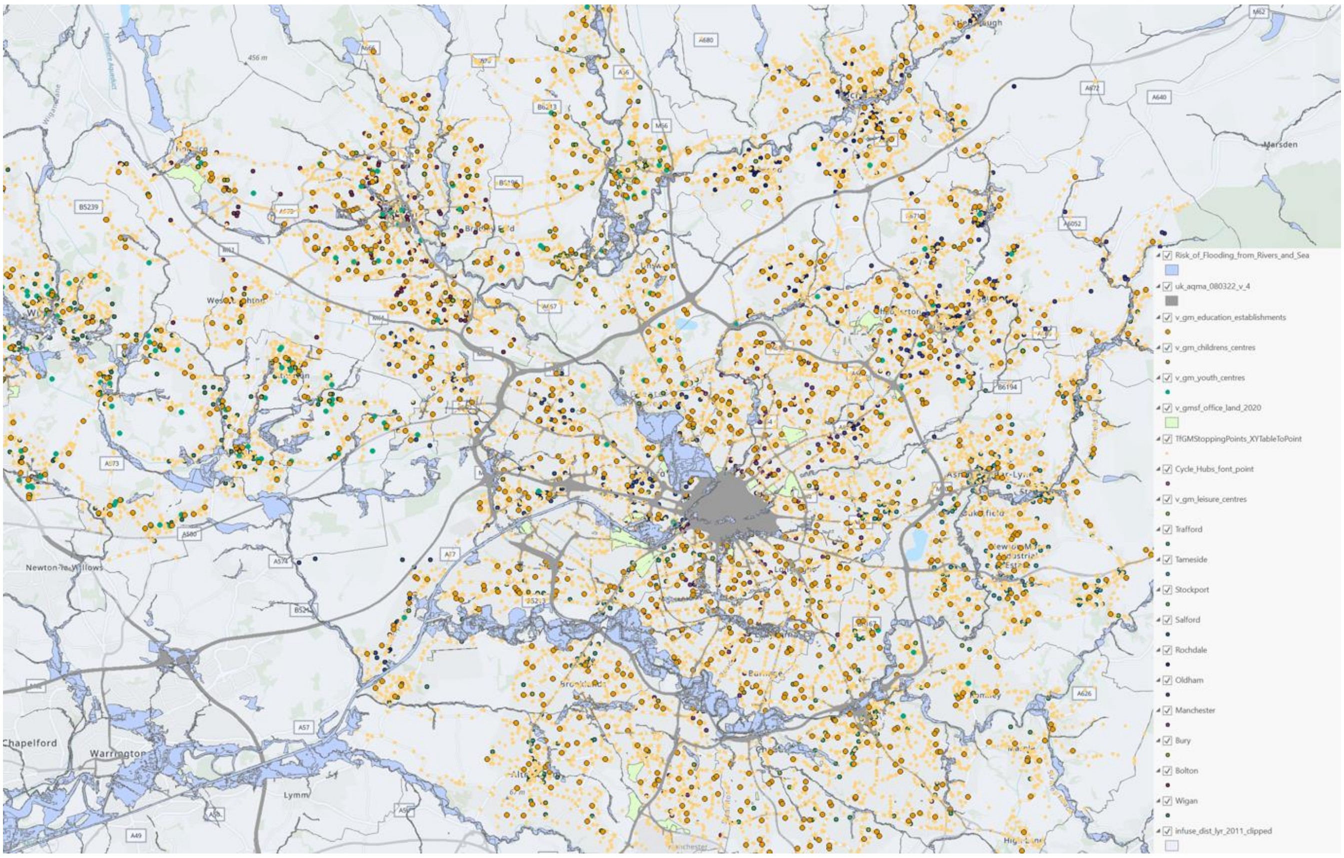

Fig 5.3 The initial map after importing all datasets.

Next, open the 'Brownfield_analysis' model builder in ADS_Brownfield.tbx and import it into ArcGIS and click the Run button to get the final Housing Land Map. If you wish to change the parameters, please find the details of each tool in the Model Builder in the previous section and change the parameters in the corresponding tool's parameters screen. For example, if you right click on Raster Calculator, you will see the corresponding parameter screen where you can change the weight of the corresponding factor, as shown in Fig 5.4.

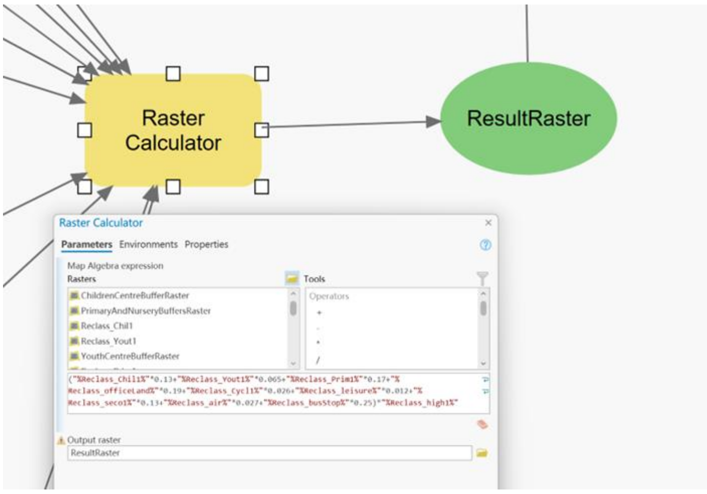

Fig 5.4 Raster Calculator.

5.2 Housing land map

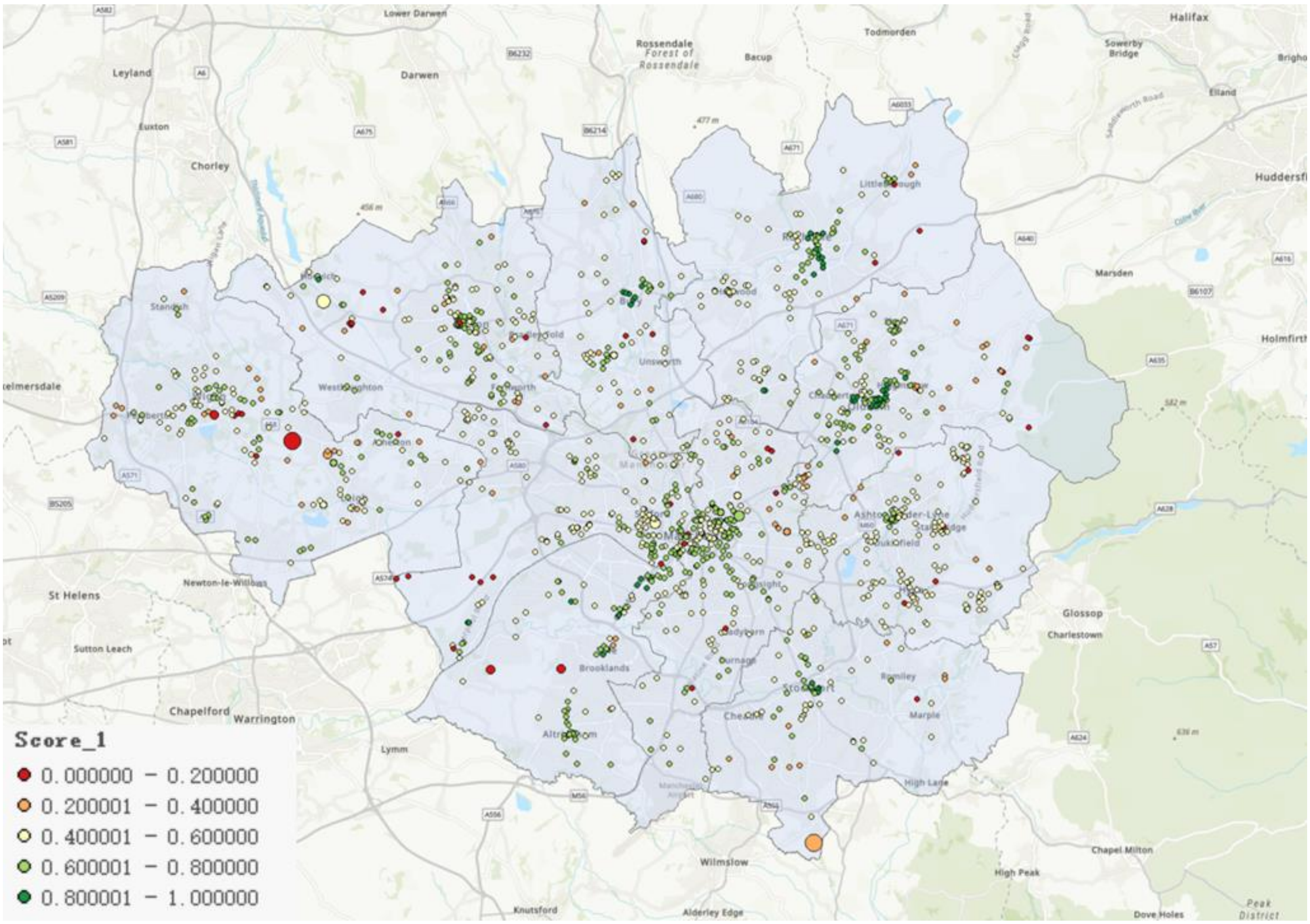

Fig 5.5 Final score of brownfields

After running the Model Builder, the Brownfield_merge layer will have an additional attribute: score, which is the score for whether brownfield is suitable for building houses. As shown in **Fig 5.5**, each point represents a brown field plot, and the size of the point represents the area of the plot. The color of dark red indicates that the brown field's score is less than 0.2, indicating that there are few service facilities around the plot, or it is in the high-risk area of flood disaster, so it is not recommended to develop as housing land; green points represent a score of 0.6 or higher, of which dark green points have a score of 0.8 or higher, that is, it is recommended to give priority to development as housing land. From the distribution point of view, there are more light green points in the middle area of Greater Manchester, and the two local authorities in the northeast have an obvious cluster of brown fields with a score greater than 0.8, and these plots can be considered first. Users can also click on specific points on the map to obtain specific information about the plot and conduct in-depth research. Of course, users can also build their own models by modifying the parameters and running logic of Model Builder, and then run Model Builder to generate results again.

6. CONCLUSION

6.1 Progress and difficulties

By preparing the Applied Data Science project through ArcGIS Pro, we gained experience and expertise on the methodologies involved in this project. We employed Python functions and imported many external factors using the ArcPy library and map manipulation technologies, resulting in a better selection of Greater Manchester housing land. Better visualisation and analytical techniques provided by the ArcGIS Pro helped us in utilising the map based geological features without any sophisticated measures. We learnt certain practical concepts in the tool like GeoX and GeoY (longitudinal and latitudinal), data conversion from CSV to SHP files, integrating ArcPy maps with the help of ArcPy package, importing SHP features to the current map, using Model Builder to create analysis blocks on the Greater Manchester important regions for consideration and using appropriate Coordinate system such as WGS84 CS. Because of the given data's different coordinate systems and the availability of current data, we learned the importance of using the most up-to-date set of data and tracking data changes through GeoServer and official government websites. We also discovered that for the Geolocation reference, the spatial reference as well as the coordinate system was considered, hence the features like SHP files needed to possess the same coordinate system otherwise there was a problem of incompatibility with the ArcPy map, which further led to incorrect plotting of those points on the map.

For the challenges part, the first issue for us is that only one member in our group had geographic knowledge, so other remaining members needed to learn the basic knowledge first, such as coordinate reference system, ArcGIS pro, etc. Another challenge was the usage of crawl method to obtain the latest datasets from official websites. Even the realization of the conversion method of different coordinate reference system and the transformer from CSV files to SHP files in Python language. At last, how to integrate the different tools to achieve a complete automated tool is exceedingly difficult without a specific reference as we also need to consider the transferability of the tool and the ease of use of the tool.

6.2 Shortcomings and further Study

This automation tool also has some shortcomings. Firstly, the stability of the tool is not sufficient. As some websites have anti-crawler measures and update their data links every now and then (e.g., flooding risk data links). The tool cannot guarantee that all data will be downloaded successfully. Therefore, it is sometimes necessary for the user to download the data that has not been downloaded. The second point is that all work is done on the local computer, which makes it easy for the user to modify the parameters and weights to meet the requirements. However, when there is a large amount of data, the performance of the computer is very demanding and if the computer is not powerful enough, the data will be processed very slowly, especially for certain steps in the Model Builder, such as 'clip' or 'buffer '.

If we were more skilled in ArcGIS online-related technology and had more time, we think it would be a better way to do all the work through the online server. For example, ArcGIS provides an online server for users to manipulate the data and analyse it through ArcGIS online, and then present the final map to the user, which might be a way to optimise the process of the tool. This approach would be more appropriate for larger scale projects, and using this approach, interaction with the user could also be achieved.

6.3 How to adapt to a larger scale

One of the future applications of this tool stands to be what if the scope of the project was extended to whole UK. There had to be a lot of changes to be made and new challenges would soon emerge. First, the domain of the data needs to be changed and we need to consider features files for whole of UK. With the increase in the project's domain there will be increase in the size of the files as well as the execution time. Secondly, re-designing of the Model Builders as well as the change in data pre-processing structure will have to be considered. For Greater Manchester, the analysis part is based on the Brownfield and other councils but for whole UK, we need to collect all the factors and data based on development standards put out by different councils. More time will be needed for carefully categorising each domain's data, hence one of the major challenges would be to determine a set of unified analysis standards i.e., the parameters used in the tool's Model Builder. As the AHP method is being used in determining weights for each factor, much bigger region would increase the workload for the overall weight system hence would increase overall execution time. Data content and format also need to be unified hence handling of such inconsistencies will be a big problem and would require a bigger group of teammates for the project.

6.4 The meaning of our tool

Despite all these above stated challenges and difficulties, this project for finding more land as well as a better land in Greater Manchester for housing needs, will contribute to the society as a complete solution to the problem most of the consumers are facing. The analysed portions of the Greater Manchester will show all the factors such as bus station, education, youth centres, leisure centres, cycle hubs, air pollution, risk of flooding from river and sea, children centre, office land, etc. These factors will play a key role in the decision of buying a land or a plot with the above-mentioned benefits and danger markers. This model can be used to purchase lands as well as lending them to customers. Our model of plot can also be used in various sectors like - Environmental services, localities, mining regions sites, Utilities, Transportation sites, Surveying tools for plots with benefits, etc. It would allow customers in improving business decisions as well as improving overall business performance and reducing cost of the plot too based on the several measured factors. It would also be beneficial for various locality services where sellers of these regions can enhance customer interaction and service as well as increase their overall profits in sales. By putting the power of spatial analysis of the plots in the hands of customers, this model will be effective in increasing overall awareness of features and factors that can lead to a better plot purchase and settlement as well.

REFERENCE

[1] "What are Greenfield and Brownfield Sites? | Slater + Gordon", 2022. [Online]. Available: https://www.slatergordon.co.uk/newsroom/greenfield-brownfield-sites/. [Accessed: 03- May- 2022].

[2] "Brownfield land registers", *GOV.UK*, 2022. [Online]. Available: https://www.gov.uk/guidance/brownfield-land-registers. [Accessed: 03- May- 2022].

[3] 2021. 20-Minute Neighbourhoods – Creating Healthier, Active, Prosperous Communities. [ebook] TCPA. Available at: <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://tcpa.org.uk/wp-content/uploads/2021/11/final_20mnguide-compressed.pdf> [Accessed 18 April 2022].

[4] "Quickstart — Requests 2.27.1 documentation", *Docs.python-requests.org*, 2022. [Online]. Available: https://docs.python-requests.org/en/latest/user/quickstart/. [Accessed: 11- May- 2022].

[5] "Beautiful Soup Documentation — Beautiful Soup 4.9.0 documentation", *Crummy.com*, 2022. [Online]. Available: https://www.crummy.com/software/BeautifulSoup/bs4/doc/. [Accessed: 11- May- 2022].

[6] "About - GeoServer", *Geoserver.org*, 2022. [Online]. Available: https://geoserver.org/about/. [Accessed: 11- May- 2022].

[7] Ordnance Survey. "A guide to coordinate systems in Great Britain: An introduction to mapping coordinate systems and the use of GPS datasets with Ordnance Survey mapping." (2015).

[8] "Transformer — pyproj 3.3.1 documentation", *Pyproj4.github.io*, 2022. [Online]. Available: https://pyproj4.github.io/pyproj/stable/api/transformer.html?highlight=transformer%20from_crs#pyproj.transformer.Transformer.from_crs. [Accessed: 11- May- 2022].

[9] "geopandas.GeoDataFrame — GeoPandas documentation", *Geopandas.org*, 2022. [Online]. Available: https://geopandas.org/en/stable/docs/reference/api/geopandas.GeoDataFrame.html. [Accessed: 11- May- 2022].

[10] "Tutorial: Getting started with arcpy.mp—ArcGIS Pro | Documentation", *Pro.arcgis.com*, 2022. [Online]. Available: https://pro.arcgis.com/en/pro-app/2.8/arcpy/mapping/tutorial-getting-started-with-arcpy-mp.htm. [Accessed: 11- May- 2022].

[11] V. Rajput and D. Dr.A.C.Shukla, "Decision-Making using the Analytic Hierarchy Process (AHP)", *International Journal of Scientific Research*, vol. 3, no. 6, pp. 135-136, 2012. Available: 10.15373/22778179/june2014/47.

APPENDIX1 Links of some useful tool

1. **AHP calculator:** https://bpmsg.com/ahp/ahp-calc.php
2. **Find the header of your chrome:** https://java2blog.com/how-to-view-http-headers-google-chrome/
3. **Run Python Code:** https://pro.arcgis.com/en/pro-app/2.8/arcpy/get-started/installing-python-for-arcgis-pro.htm
4. **The link of our code and Arc Tool box**: https://github.com/MascaraLiz/Housing-land.git

APPENDIX 2 Model Builder