Workshop 4 - Wolverhampton Hierarchical clustering

Student Number: 2302546

In [45]:
```
import pandas as pd
import matplotlib.pyplot as plt
import folium
import os, re
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import normalize
from IPython.display import IFrame
from sklearn.cluster import AgglomerativeClustering
import scipy.cluster.hierarchy as shc

%matplotlib inline

import warnings
warnings.filterwarnings('ignore')
```

For this notebook, I am considering a period of December 2021 to December 2022

In [2]:
```
path_to_data = './crime'
cd = os.path.dirname(os.path.abspath(path_to_data))

i = 0
columns = range(1,100)
dfList = []

for root, dirs, files in os.walk(cd):
    for fname in files:
        if re.match("^.*.csv$", fname):
            frame = pd.read_csv(os.path.join(root, fname))
            frame['key'] = "file{}".format(i)
            dfList.append(frame)
            i += 1

dataset = pd.concat(dfList)
```

In [3]:
```
dataset.head()
```

Out[3]:

	Crime ID	Month	Reported by	Falls within	Longitude	L
0	NaN	2021-12	West Midlands Police	West Midlands Police	-1.850772	52
1	01988dde64ab85a563fdbf7b7d48bb1c64163446defb36...	2021-12	West Midlands Police	West Midlands Police	-1.851382	52
2	80a7628f3737fe0f5fe1d3835f44730498309bb28565f2...	2021-12	West Midlands Police	West Midlands Police	-1.851382	52
3	5d2ea684f34a8c6ffd574c8a61b326447a3c9384174084...	2021-12	West Midlands Police	West Midlands Police	-1.849280	52
4	552eac09561aa50f4c68db0a4c0d11512c80066ea2f55a...	2021-12	West Midlands Police	West Midlands Police	-1.840641	52

In [4]:
```
print(dataset.shape)
```
```
(2510292, 13)
```

In [5]:
```
name_number = 'AkinyemiArabambi-2302546.csv'
dataset.to_csv(name_number, index=False)
```

In [6]:
```
data = pd.read_csv(name_number)
```

In [7]:
```
data['Crime type'].value_counts()
```

Out[7]:
```
Violence and sexual offences    1073298
Vehicle crime                    244020
Public order                     217134
Criminal damage and arson        192186
Other theft                      167760
Burglary                         138522
Anti-social behaviour            135048
Shoplifting                      105486
Robbery                           53544
Drugs                             51480
Possession of weapons             45744
Other crime                       41364
Theft from the person             29004
Bicycle theft                     15702
Name: Crime type, dtype: int64
```

Q1. Use a similar approach to display number of crimes in each month. You can

use the "Month" column to do that.

In [8]:
```
data['Month'].value_counts()
```

Out[8]:
```
2022-07    207492
2022-05    204630
2022-08    202050
2022-06    200004
2022-10    197748
2022-03    195582
2021-12    193224
2022-04    191058
2022-09    190182
2022-01    187698
2022-11    185784
2022-02    178782
2022-12    176058
Name: Month, dtype: int64
```

In [9]:
```
data['town'] = data['LSOA name'].str.split(' ').str[0]
```

In [10]:
```
data.head()
```

Out[10]:

	Crime ID	Month	Reported by	Falls within	Longitude	L
0	NaN	2021-12	West Midlands Police	West Midlands Police	-1.850772	52
1	01988dde64ab85a563fdbf7b7d48bb1c64163446defb36...	2021-12	West Midlands Police	West Midlands Police	-1.851382	52
2	80a7628f3737fe0f5fe1d3835f44730498309bb28565f2...	2021-12	West Midlands Police	West Midlands Police	-1.851382	52
3	5d2ea684f34a8c6ffd574c8a61b326447a3c9384174084...	2021-12	West Midlands Police	West Midlands Police	-1.849280	52
4	552eac09561aa50f4c68db0a4c0d11512c80066ea2f55a...	2021-12	West Midlands Police	West Midlands Police	-1.840641	52

In [11]:
```
towns = ['Wolverhampton']
filtered_data = data[data.town.str.contains('|'.join(towns), na=False)]
filtered_data.head()
```

Out[11]:

	Crime ID	Month	Reported by	Falls within	Longitud
29180	NaN	2021-12	West Midlands Police	West Midlands Police	-2.12917
29181	8c4d1389c9f1264729c9c2ca5ed47d759cc56fb37dffb7...	2021-12	West Midlands Police	West Midlands Police	-2.11996
29182	8e826565d9c135877029412e45c3dc682441c28250ad5d...	2021-12	West Midlands Police	West Midlands Police	-2.12638
29183	ca6291400db992354fbf9a445c9973587dd629a91c1ef3...	2021-12	West Midlands Police	West Midlands Police	-2.12638
29184	13ae1c0077256e3f81e3f0fdca8bf70e3f4f5c44fa5dc9...	2021-12	West Midlands Police	West Midlands Police	-2.12970

Q2. Display crime types in Wolverhampton.

In [12]:

```
print('The crime types in Wolverhampton are: ', filtered_data['Crime type'].unique
```

```
The crime types in Wolverhampton are:  ['Anti-social behaviour' 'Burglary' 'Drugs'
'Public order' 'Robbery'
 'Shoplifting' 'Vehicle crime' 'Violence and sexual offences'
 'Other theft' 'Criminal damage and arson' 'Possession of weapons'
 'Other crime' 'Theft from the person' 'Bicycle theft']
```

Q3. What is the most common crime committed in Wolverhampton in the dataset? Is it the same most common crime in West Midlands too?

In [13]:

```
filtered_data['Crime type'].value_counts()
```

Out[13]:

```
Violence and sexual offences    111324
Public order                     21822
Vehicle crime                    19122
Criminal damage and arson        18768
Other theft                      14676
Burglary                         13590
Anti-social behaviour            11274
Shoplifting                      10848
Other crime                       4488
Drugs                             4404
Robbery                           4110
Possession of weapons             3876
Theft from the person             1878
Bicycle theft                     1644
Name: Crime type, dtype: int64
```

In [14]:
```
filtered_data['LSOA code'].value_counts().nlargest(10)
```

Out[14]:
```
E01010521    23484
E01010564     8100
E01010414     5094
E01010530     4386
E01010410     4170
E01010450     3846
E01010472     3306
E01010473     3186
E01010508     3144
E01010476     3090
Name: LSOA code, dtype: int64
```

Q4. Provide a prime landmark of minimum of 2 LSOA codes under investigation. If there is no recognisable prime landmark, provide name(s) of the nearby streets/roads surrounding those areas.

LSOA code E01010521 which is the Wolverhampton City Center which includes the area around the University of Wolverhampton has the highest crime rate of 23484 crimes in Wolverhampton (https://www.doogal.co.uk/LSOA?code=E01010521)

LSOA code E01010564 is the area enclosed in some parts of Lichfield Road, Lakefield Rd, Watery Ln and Wednesfield Way has the second highest crime rate of 8100 in the Wolverhampotn region (https://www.doogal.co.uk/LSOA?code=E01010564)

In [15]:
```
filtered_important_data = filtered_data[['LSOA code','Crime type']]
filtered_important_data = pd.get_dummies(filtered_important_data, columns=['Crime 
clustering_data = filtered_important_data.groupby(['LSOA code']).agg(
    {'Crime type_Anti-social behaviour':'sum',
     'Crime type_Bicycle theft':'sum',
     'Crime type_Burglary':'sum',
     'Crime type_Criminal damage and arson':'sum',
     'Crime type_Drugs':'sum',
     'Crime type_Other crime':'sum',
     'Crime type_Other theft':'sum',
     'Crime type_Possession of weapons':'sum',
     'Crime type_Public order':'sum',
     'Crime type_Robbery':'sum',
     'Crime type_Shoplifting':'sum',
     'Crime type_Theft from the person':'sum',
     'Crime type_Vehicle crime':'sum',
     'Crime type_Violence and sexual offences':'sum'
    }
    ).reset_index()
```

In [16]:
```
clustering_data[:5]
```

Out[16]:

	LSOA code	Crime type_Anti-social behaviour	Crime type_Bicycle theft	Crime type_Burglary	Crime type_Criminal damage and arson	Crime type_Drugs	Crime type_Other crime	typ
0	E01010410	222.0	42.0	204.0	318.0	18.0	60	
1	E01010411	30.0	6.0	30.0	144.0	6.0	24	
2	E01010412	60.0	0.0	60.0	144.0	24.0	78	
3	E01010413	54.0	24.0	66.0	186.0	24.0	30	
4	E01010414	258.0	48.0	396.0	522.0	48.0	72	

In [17]:
```
clustering_data_original = clustering_data.copy()
clustering_data_original.head()
```

Out[17]:

	LSOA code	Crime type_Anti-social behaviour	Crime type_Bicycle theft	Crime type_Burglary	Crime type_Criminal damage and arson	Crime type_Drugs	Crime type_Other crime	typ
0	E01010410	222.0	42.0	204.0	318.0	18.0	60	
1	E01010411	30.0	6.0	30.0	144.0	6.0	24	
2	E01010412	60.0	0.0	60.0	144.0	24.0	78	
3	E01010413	54.0	24.0	66.0	186.0	24.0	30	
4	E01010414	258.0	48.0	396.0	522.0	48.0	72	

In [18]:
```
clustering_data.drop(['LSOA code'], axis = 1, inplace = True, errors = 'ignore')
clustering_data.head()
```

Out[18]:

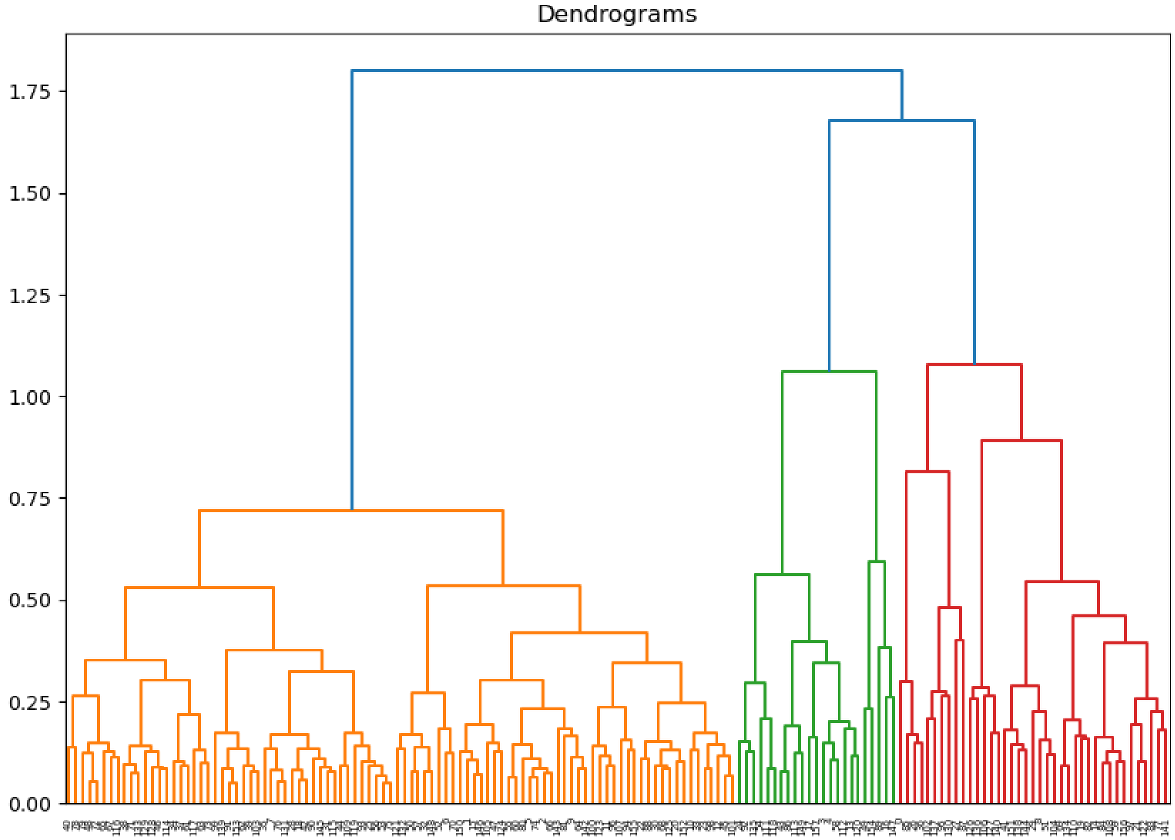
	Crime type_Anti-social behaviour	Crime type_Bicycle theft	Crime type_Burglary	Crime type_Criminal damage and arson	Crime type_Drugs	Crime type_Other crime	Crime type_Other theft	typ
0	222.0	42.0	204.0	318.0	18.0	60	876.0	
1	30.0	6.0	30.0	144.0	6.0	24	36.0	
2	60.0	0.0	60.0	144.0	24.0	78	72.0	
3	54.0	24.0	66.0	186.0	24.0	30	72.0	
4	258.0	48.0	396.0	522.0	48.0	72	204.0	

In [19]:
```
data_scaled = normalize(clustering_data)
data_scaled = pd.DataFrame(data_scaled, columns=clustering_data.columns)
data_scaled.head()
```

Out[19]:

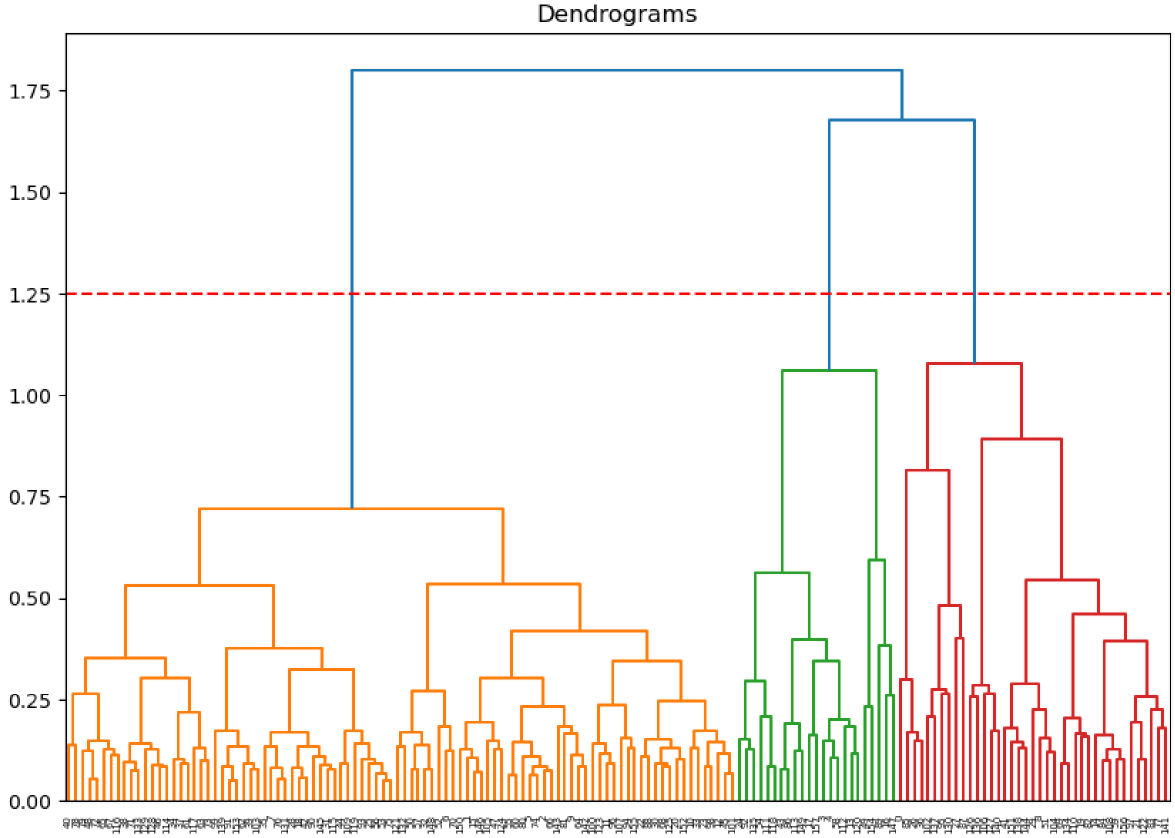
	Crime type_Anti-social behaviour	Crime type_Bicycle theft	Crime type_Burglary	Crime type_Criminal damage and arson	Crime type_Drugs	Crime type_Other crime	Crime type_Other theft	typ
0	0.129000	0.024405	0.118540	0.184783	0.010459	0.034865	0.509026	
1	0.049600	0.009920	0.049600	0.238079	0.009920	0.039680	0.059520	
2	0.058126	0.000000	0.058126	0.139502	0.023250	0.075564	0.069751	
3	0.067071	0.029809	0.081975	0.231022	0.029809	0.037262	0.089428	
4	0.117179	0.021801	0.179857	0.237084	0.021801	0.032701	0.092653	

In [20]:
```
plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = shc.dendrogram(shc.linkage(data_scaled, method='ward'))
```



In [21]:
```
plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = shc.dendrogram(shc.linkage(data_scaled, method='ward'))
plt.axhline(y=1.25, color='r', linestyle='--')
```

Out[21]: <matplotlib.lines.Line2D at 0x190197d14f0>



Q5. Discuss what happens when you decide to cut the dendogram at a different level.

Cutting a dendrogram at a different level means selecting a different threshold value for the height at which to "cut" the dendrogram to obtain a set number of clusters. When cutting the dendrogram at a lower level in this instance at 1.0 will result in 5 clusters as compared to the current 3 clusters.

If the dendrogram is cut at a high level, the resulting clusters will be more general and inclusive, meaning that similar data points will be grouped together regardless of their finer differences. This may be appropriate for exploratory data analysis or for identifying broad patterns in the data.

On the other hand, cutting the dendrogram at a low level may result in more specific and granular clusters, which can reveal more subtle patterns in the data. However, if the dendrogram is cut too low, it may result in overfitting or generating clusters that are too specific to the particular data at hand, leading to poor generalization and limited usefulness for future analyses.

Therefore, the decision of where to cut the dendrogram should be made carefully and should consider the specific goals of the analysis and the nature of the data being clustered. It is often helpful to explore the results of clustering at different levels to gain a better understanding of the structure of the data and to choose a suitable level for the final set of clusters.

In [22]:
```
cluster = AgglomerativeClustering(n_clusters=3, affinity='euclidean', linkage='ward
cluster_ids = cluster.fit_predict(data_scaled)
```

In [23]:
```
clustering_data['cluster'] = cluster_ids
clustering_data.head()
```

Out[23]:

	Crime type_Anti-social behaviour	Crime type_Bicycle theft	Crime type_Burglary	Crime type_Criminal damage and arson	Crime type_Drugs	Crime type_Other crime	Crime type_Other theft	typ
0	222.0	42.0	204.0	318.0	18.0	60	876.0	
1	30.0	6.0	30.0	144.0	6.0	24	36.0	
2	60.0	0.0	60.0	144.0	24.0	78	72.0	
3	54.0	24.0	66.0	186.0	24.0	30	72.0	
4	258.0	48.0	396.0	522.0	48.0	72	204.0	

In [24]:
```
hierarchical_cluster = pd.DataFrame(round(clustering_data.groupby('cluster').mean(
hierarchical_cluster
```

Out[24]:

	Crime type_Anti-social behaviour	Crime type_Bicycle theft	Crime type_Burglary	Crime type_Criminal damage and arson	Crime type_Drugs	Crime type_Other crime	Crim type_Othe thef
cluster							
0	32.3	4.2	56.8	55.2	9.8	18.0	81.4
1	70.1	7.3	78.6	123.1	24.1	29.8	72.2
2	143.0	33.9	166.7	208.7	74.3	40.2	198.8

Q6. Discuss the clustering results based on your dataset.

Based on observation, the clustering results can be observed to show that:

- Cluster ID 2 is the LSOA codes with the highest number of crimes (the highest risk regions).
- Cluster ID 0 is the LSOA codes with the lowest number of crimes (the lowest risk regions).
- Cluster ID 1 is the LSOA codes with moderate number of crimes.

In [25]:
```
clustering_data_original['cluster'] = cluster_ids
clusters = clustering_data_original[['LSOA code', 'cluster']]
```

In [26]:
```
clusters.head()
```

Out[26]:

	LSOA code	cluster
0	E01010410	0
1	E01010411	1
2	E01010412	1
3	E01010413	2
4	E01010414	2

In [27]:
```
clusters.shape
```

Out[27]:
```
(158, 2)
```

In [28]:
```
clustered_full = pd.merge(filtered_data, clusters, on='LSOA code')
clustered_full.head()
```

Out[28]:

	Crime ID	Month	Reported by	Falls within	Longitude	l
0	NaN	2021-12	West Midlands Police	West Midlands Police	-2.129177	52
1	8c4d1389c9f1264729c9c2ca5ed47d759cc56fb37dffb7...	2021-12	West Midlands Police	West Midlands Police	-2.119967	52
2	8e826565d9c135877029412e45c3dc682441c28250ad5d...	2021-12	West Midlands Police	West Midlands Police	-2.126382	52
3	ca6291400db992354fbf9a445c9973587dd629a91c1ef3...	2021-12	West Midlands Police	West Midlands Police	-2.126382	52
4	13ae1c0077256e3f81e3f0fdca8bf70e3f4f5c44fa5dc9...	2021-12	West Midlands Police	West Midlands Police	-2.129703	52

In [29]:
```
def get_color(cluster_id):
    if cluster_id == 2:
        return 'darkred'
    if cluster_id == 0:
        return 'green'
    if cluster_id == 1:
        return 'amber'
```

In [30]:
```
#create a map
this_map = folium.Map(location =[clustered_full["Latitude"].mean(),
                                  clustered_full["Longitude"].mean()], zoom_start=5

def plot_dot(point):
    '''input: series that contains a numeric named latitude and a numeric named lon
    this function creates a CircleMarker and adds it to your this_map'''
    folium.CircleMarker(location=[point.Latitude, point.Longitude],
```

```
                                radius=2,
                                color=point.color,
                                weight=1).add_to(this_map)


clustered_full["color"] = clustered_full["cluster"].apply(lambda x: get_color(x))

#use df.apply(,axis=1) to iterate through every row in your dataframe
clustered_full.apply(plot_dot, axis = 1)


#Set the zoom to the maximum possible
this_map.fit_bounds(this_map.get_bounds())

#Save the map to an HTML file
this_map.save(os.path.join('Crime_map.html'))
#IFrame(src='Crime_map.html', width=1000, height=600)
```

Q7. Discuss the results (your understanding) of the clustering algorithm on clustering crime dataset used in this work.

- Wolverhampton

The clustering algorithm used in this work is agglomerative hierarchical clustering, which is a bottom-up approach that builds a hierarchy of clusters. The dendrogram is used to determine the optimal number of clusters to divide the dataset. In this dataset, the dendrogram shows that cutting it at a height of 1.25 gives a reasonable number of four clusters as the blue line had the longest vertical distance and could be cut easily at 1.25 without interference from the horizontal lines.

Based on the results of the clustering algorithm, it can be observed that the LSOA codes with the highest number of crimes (the highest risk regions) belong to Cluster ID 2. Conversely, the LSOA codes with the lowest number of crimes (the lowest risk regions) belong to Cluster ID 0. The LSOA codes with a moderate number of crimes belong to Cluster ID 1.

On the crime map, it is seen as we have color coded each cluster with darkred, green, amber and blue for CLuster 2, 0, and 1 respectively that areas around the main Wolverhampton city center has an high rate of crime as it is fully filled with amber-red color. Also several news outlet has news on the rate of crime within this time range in the Wolverhampton City Center (https://www.expressandstar.com/news/crime/2022/12/27/man-remains-in-hospital-after-city-centre-assault/)

It is essential to note that the clustering algorithm is unsupervised, and the results obtained from the clustering analysis depend on the features used in the analysis. In this work, we used the different types of crimes committed. The results can, therefore, be used to guide policy-making and resource allocation by providing information on areas where crime is most prevalent and where law enforcement agencies should concentrate their efforts.

Q8. Change the parameters of the algorithm as appropriate and perform the clustering algorithm and draw the graph again. Discuss your results briefly.

In [31]:
```
filtered_data['Month'].value_counts()
```

Out[31]:
```
2022-08    19806
2022-07    19788
2022-06    19320
2022-10    19272
2022-05    19110
2022-04    19086
2022-03    18834
2022-09    18642
2021-12    18144
2022-01    17964
2022-02    17424
2022-11    17406
2022-12    17028
Name: Month, dtype: int64
```

In [32]:
```
filtered_important_data1= filtered_data[['LSOA code','Month']]
filtered_important_data1= pd.get_dummies(filtered_important_data1,columns=['Month'
clustering_data1= filtered_important_data1.groupby(['LSOA code']).agg(
    {'Month_2022-05':'sum',
     'Month_2022-01':'sum',
     'Month_2022-03':'sum',
     'Month_2022-04':'sum',
     'Month_2022-07':'sum',
     'Month_2022-06':'sum',
     'Month_2022-02':'sum',
     'Month_2022-08':'sum',
     'Month_2022-09':'sum',
     'Month_2022-10':'sum',
     'Month_2021-12':'sum',
     'Month_2022-11':'sum',
     'Month_2022-12':'sum',
    }
    ).reset_index()
```

In [33]:
```
clustering_data1[:5]
```

Out[33]:

	LSOA code	Month_2022-05	Month_2022-01	Month_2022-03	Month_2022-04	Month_2022-07	Month_202
0	E01010410	270.0	318.0	330.0	378.0	348.0	228
1	E01010411	114.0	120.0	78.0	90.0	78.0	96
2	E01010412	162.0	120.0	120.0	168.0	222.0	252
3	E01010413	180.0	168.0	108.0	168.0	150.0	96
4	E01010414	252.0	444.0	384.0	402.0	294.0	294

In [34]:
```
clustering_data_original1 = clustering_data1.copy()
clustering_data_original1.head()
```

Out[34]:

	LSOA code	Month_2022-05	Month_2022-01	Month_2022-03	Month_2022-04	Month_2022-07	Month_202
0	E01010410	270.0	318.0	330.0	378.0	348.0	228
1	E01010411	114.0	120.0	78.0	90.0	78.0	96
2	E01010412	162.0	120.0	120.0	168.0	222.0	252
3	E01010413	180.0	168.0	108.0	168.0	150.0	96
4	E01010414	252.0	444.0	384.0	402.0	294.0	294

In [35]:

```
clustering_data1.drop(['LSOA code'], axis = 1, inplace = True, errors = 'ignore')
clustering_data1.head()
```

Out[35]:

	Month_2022-05	Month_2022-01	Month_2022-03	Month_2022-04	Month_2022-07	Month_2022-06	Month_2
0	270.0	318.0	330.0	378.0	348.0	228.0	
1	114.0	120.0	78.0	90.0	78.0	96.0	
2	162.0	120.0	120.0	168.0	222.0	252.0	
3	180.0	168.0	108.0	168.0	150.0	96.0	
4	252.0	444.0	384.0	402.0	294.0	294.0	

In [36]:

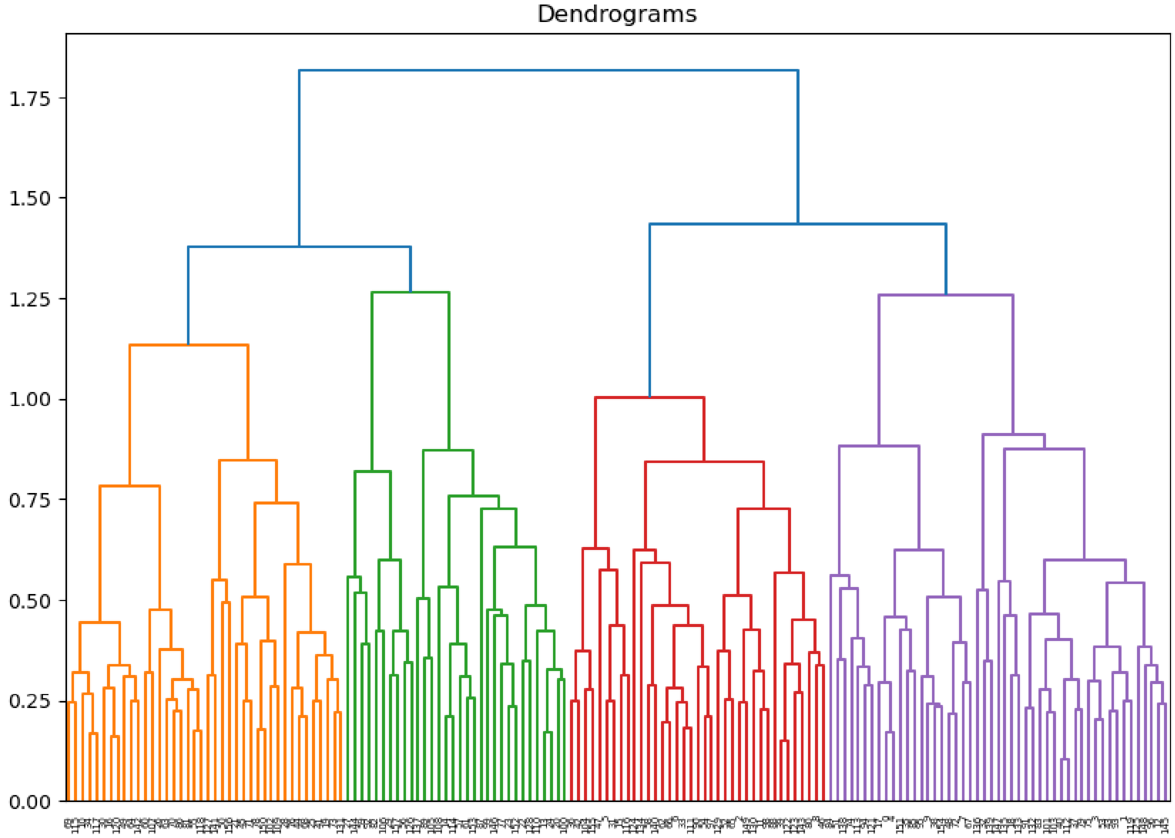
```
data_scaled1 = normalize(clustering_data1)
data_scaled1 = pd.DataFrame(data_scaled1, columns=clustering_data1.columns)
data_scaled1.head()
```

Out[36]:

	Month_2022-05	Month_2022-01	Month_2022-03	Month_2022-04	Month_2022-07	Month_2022-06	Month_2
0	0.229195	0.269941	0.280128	0.320874	0.295407	0.193543	0.2
1	0.379696	0.399680	0.259792	0.299760	0.259792	0.319744	0.1
2	0.297152	0.220113	0.220113	0.308158	0.407209	0.462237	0.2
3	0.375705	0.350658	0.225423	0.350658	0.313088	0.200376	0.2
4	0.175514	0.309239	0.267450	0.279987	0.204767	0.204767	0.3

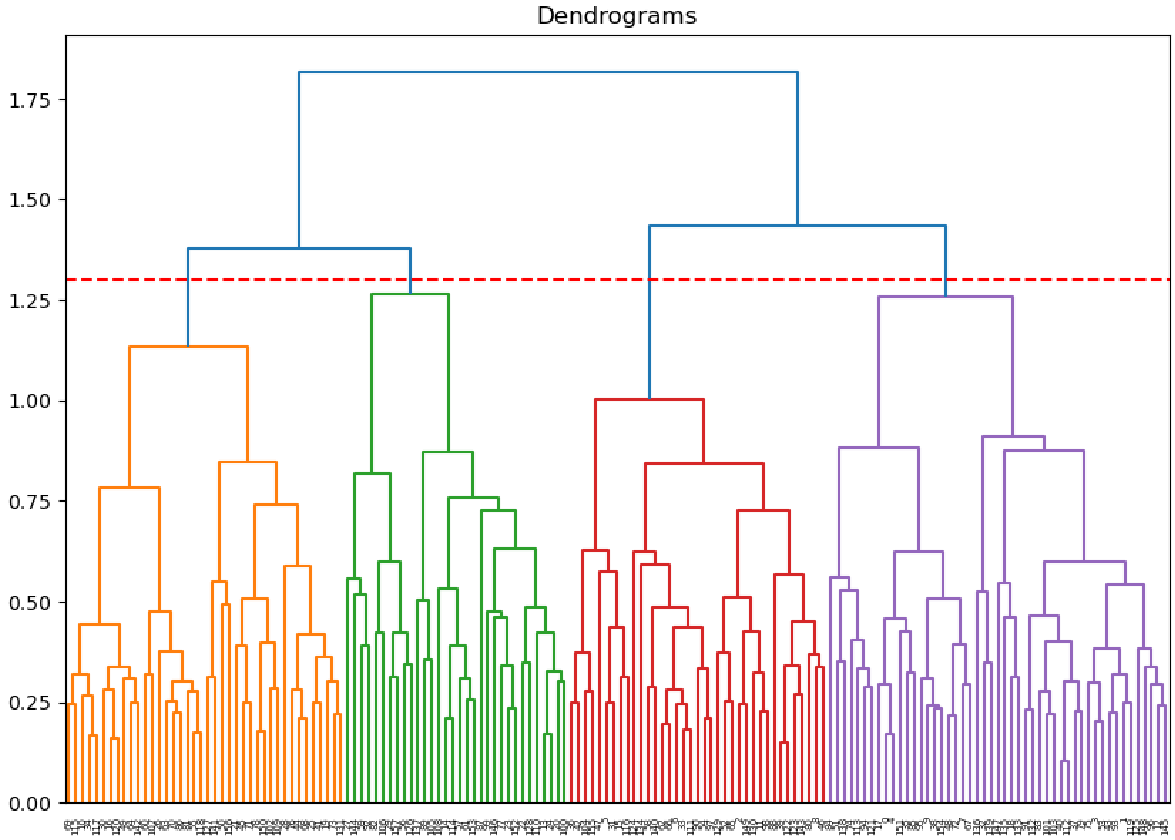
In [37]:

```
plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = shc.dendrogram(shc.linkage(data_scaled1, method='ward'))
```



In [38]:
```python
plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = shc.dendrogram(shc.linkage(data_scaled1, method='ward'))
plt.axhline(y=1.3, color='r', linestyle='--')
```

Out[38]: <matplotlib.lines.Line2D at 0x19069472e80>



In [42]:
```
cluster1 = AgglomerativeClustering(n_clusters=4, affinity='euclidean', linkage='wa
cluster_ids1 = cluster1.fit_predict(data_scaled1)
```

In [43]:
```
clustering_data1['cluster'] = cluster_ids1
clustering_data1.head()
```

Out[43]:

	Month_2022-05	Month_2022-01	Month_2022-03	Month_2022-04	Month_2022-07	Month_2022-06	Month_2
0	270.0	318.0	330.0	378.0	348.0	228.0	
1	114.0	120.0	78.0	90.0	78.0	96.0	
2	162.0	120.0	120.0	168.0	222.0	252.0	
3	180.0	168.0	108.0	168.0	150.0	96.0	
4	252.0	444.0	384.0	402.0	294.0	294.0	

In [44]:
```
hierarchical_cluster1 = pd.DataFrame(round(clustering_data1.groupby('cluster').mea
hierarchical_cluster1
```

Out[44]:

cluster	Month_2022-05	Month_2022-01	Month_2022-03	Month_2022-04	Month_2022-07	Month_2022-06	Mo
0	56.2	62.4	76.9	82.5	75.9	60.9	
1	144.4	143.3	133.7	129.2	105.8	123.3	
2	160.1	143.4	169.9	157.8	207.9	187.6	
3	107.8	91.0	88.4	107.0	112.0	109.6	

Based on observation, the clustering results can be observed to show that:

- Cluster ID 2 is the LSOA codes with the highest number of crimes (the highest risk regions).
- Cluster ID 0 is the LSOA codes with the lowest number of crimes (the lowest risk regions).
- Cluster ID 1 is the LSOA codes with moderate number of crimes.
- Cluster ID 3 is the LSOA codes where crime numbers are mild.

Q9. Consider a different towns (example : dudley) and perform the clustering again. You should choose the number of clusters from the dendrogram accordingly. Discuss your results briefly.

- Dudley

In the Dudley dataset, the dendrogram shows that cutting it at a height of 1.5 gives a reasonable number of four clusters as the blue line had the longest vertical distance and could be cut easily at 1.5 without interference from the horizontal lines.

Based on the results of the clustering algorithm, it can be observed that the LSOA codes with the highest number of crimes (the highest risk regions) belong to Cluster ID 0.

Conversely, the LSOA codes with the lowest number of crimes (the lowest risk regions) belong to Cluster ID 3. The LSOA codes with a moderate number of crimes belong to Cluster ID 1, while Cluster ID 2 contains LSOA codes where crime numbers are mild.

On the crime map, it is seen as we have color coded each cluster with darkred, green, amber and blue for CLuster 0, 3, 1 and 2 respectively that areas around the Stourbridge has an high rate of crime as it is fully filled with amber red color. The safest area with low crime rates are Wordsley and Pensnett Trading Estate Also several news outlet has news on the rate of crime within this time range in the Stourbridge (https://www.expressandstar.com/news/crime/2022/12/13/police-destroy-deadly-knife-which-owner-had-hidden-in-their-sock/)

It is essential to note that the clustering algorithm is unsupervised, and the results obtained from the clustering analysis depend on the features used in the analysis. In this work, we used the different types of crimes committed. The results can, therefore, be used to guide policy-making and resource allocation by providing information on areas where crime is most prevalent and where law enforcement agencies should concentrate their efforts.

- Wolverhampton vs Dudley

Both the Wolverhampton and Dudley datasets used agglomerative hierarchical clustering and were cut at 1.25 and 1.5 to obtain four clusters. However, the LSOA codes in the two cities were grouped differently into the four clusters based on their crime rates.

In Wolverhampton, the LSOA codes with the highest number of crimes were grouped into Cluster ID 0, and this cluster was the largest in terms of the number of LSOA codes it contained. In contrast, in Dudley, the LSOA codes with the highest number of crimes were grouped into Cluster ID 0, but this cluster was the smallest in terms of the number of LSOA codes it contained.

Similarly, in Wolverhampton, the LSOA codes with the lowest number of crimes were grouped into Cluster ID 2, while in Dudley, they were grouped into Cluster ID 3. The LSOA codes with moderate crime rates were also grouped differently in the two cities, with Wolverhampton having them in Cluster ID 3 and Dudley having them in Cluster ID 1.

On the crime maps, the areas with high crime rates were different in the two cities. In Wolverhampton, the highest rates were concentrated around the city center, while in Dudley, they were concentrated around Stourbridge. Both cities had areas with low crime rates, but they were different locations.

References

Baqir, A., ul Rehman, S., Malik, S., ul Mustafa, F. and Ahmad, U., 2020, January. Evaluating the performance of hierarchical clustering algorithms to detect spatio-temporal crime hot-spots. In 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-5). IEEE.

Sivaranjani, S., Sivakumari, S. and Aasha, M., 2016, October. Crime prediction and forecasting in Tamilnadu using clustering approaches. In 2016 International Conference on Emerging Technological Trends (ICETT) (pp. 1-6). IEEE.