Project 3 Github Modified

April 16, 2023

0.1 Imports

[97]:
```
import pandas as pd
```

0.1.1 2.1 Trees

[98]:
```
Pandas dataframe called "trees" that contains the contents of the Excel file
trees = pd.read_excel("camden_trees.xlsx")
```

[99]:
```
# Displaying the first few rows
trees.head()
```

[99]:

	Identifier	Number Of Trees	Site Name	Contract Area	Scientific Name	Inspection Date	Inspection Due Date	Height In Metres	Spread In Metres	Diameter In Centimetres At Breast Height	Ward Code	Ward Name	Easting	Northing	Longitude	Latitude
0	00060053	1.0	Russell Nurseries Estate	Housing	Vacant Tree Pit	NaT	NaN	NaN	NaN	NaN	E05000135	Hampstead Town	527305	185240	-0.165240	51.551693
1	00057855	1.0	BRECKNOCK JMI (E)	Education	Vacant Tree Pit	2019-07-17	2022/2023	NaN	NaN	NaN	E05000131	Cantelowes	529923	184782	-0.127681	51.546984
2	00059953	1.0	Estate 51 Ravenshaw Street	Housing	Ficus carica	NaT	NaN	5.0	4.0	10.0	NaN	NaN	0	0	NaN	NaN
3	00059915	1.0	ROSARY RC JMI (E)	Education	Betula jacquemontii	NaT	NaN	4.0	1.0	6.0	E05000135	Hampstead Town	527249	185261	-0.166051	51.551901
4	00010762	1.0	Holly Lodge Estate	Housing	Ilex x altaclarensis	2017-06-14	2020/2021	14.0	6.0	26.0	E05000137					

```
4        Highgate   528414    186770  -0.148704  51.565198

                 Location
0   (51.551693, -0.16524)
1  (51.546984, -0.127681)
2                     NaN
3  (51.551901, -0.166051)
4  (51.565198, -0.148704)
```

[100]:
```
# Displaying the number of rows and columns
trees.shape
```

[100]:
```
(23444, 17)
```

[101]:
```
# Displaying a list of all the columns in the dataframe
trees.columns
```

[101]:
```
Index(['Identifier', 'Number Of Trees', 'Site Name', 'Contract Area',
       'Scientific Name', 'Inspection Date', 'Inspection Due Date',
       'Height In Metres', 'Spread In Metres',
       'Diameter In Centimetres At Breast Height', 'Ward Code', 'Ward Name',
       'Easting', 'Northing', 'Longitude', 'Latitude', 'Location'],
      dtype='object')
```

[102]:
```
# Displaying a List the data types of each column
trees.dtypes
```

[102]:
```
Identifier                                    object
Number Of Trees                              float64
Site Name                                     object
Contract Area                                 object
Scientific Name                               object
Inspection Date                       datetime64[ns]
Inspection Due Date                           object
Height In Metres                             float64
Spread In Metres                             float64
Diameter In Centimetres At Breast Height     float64
Ward Code                                     object
Ward Name                                     object
Easting                                        int64
Northing                                       int64
Longitude                                    float64
Latitude                                     float64
Location                                      object
dtype: object
```

0.1.2 2.2 Environmental

[103]:
```
# Created a Pandas dataframe called "environmental" that contains the contents␣
↪of the csv file
environmental = pd.read_csv("camden_trees_environmental.csv")
```

[104]:
```
# Displayed the first few rows
environmental.head()
```

[104]:

	Identifier	Maturity	Physiological Condition	Tree Set To Be Removed	Removal Reason	Capital Asset Value For Amenity Trees	Carbon Storage In Kilograms	Gross Carbon Sequestration Per Year In Kilograms	Pollution Removal Per Year In Grams
0	00055125	Juvenile	Good	No	NaN	115.07	1.6	0.5	5.7
1	00059429	Middle aged	Fair	No	NaN	7518.08	NaN	NaN	NaN
2	00018254	Mature	Fair	No	NaN	20419.63	426.4	8.8	215.2
3	00027155	Mature	Fair	No	NaN	21447.74	448.3	9.6	379.1
4	00041326	Juvenile	Good	No	NaN	524.30	9.9	1.4	12.8

[105]:
```
# Displayed the number of rows and columns
environmental.shape
```

[105]:
```
(23415, 9)
```

[106]:
```
#Displayed a list of all the columns in the dataframe
environmental.columns
```

[106]:
```
Index(['Identifier', 'Maturity', 'Physiological Condition',
       'Tree Set To Be Removed', 'Removal Reason',
       'Capital Asset Value For Amenity Trees', 'Carbon Storage In Kilograms',
       'Gross Carbon Sequestration Per Year In Kilograms',
       'Pollution Removal Per Year In Grams'],
      dtype='object')
```

[107]:
```
# Listed the data types of each column
environmental.dtypes
```

[107]:
```
Identifier                                          object
Maturity                                            object
Physiological Condition                             object
Tree Set To Be Removed                              object
Removal Reason                                      object
Capital Asset Value For Amenity Trees              float64
Carbon Storage In Kilograms                        float64
Gross Carbon Sequestration Per Year In Kilograms    float64
Pollution Removal Per Year In Grams                float64
dtype: object
```

0.1.3 2.3 Common and Scientific Names

[108]:
```
# Created a Pandas dataframe called "names" that contains the contents of the␣
 ↪json file
names = pd.read_json("https://s3.eu-west-1.amazonaws.com/course.oc-static.com/
 ↪projects/DAN_UK_App_P3/tree_common_names.json")
```

[109]:
```
# Displayed the first few rows
names.head()
```

[109]:

	Scientific Name	Common Name
0	Carpinus betulus Lucas	Hornbeam - European
1	Prunus 'Pandora'	Cherry - Ornamental
2	Tilia unidentified species	Lime
3	Rosa unidentified species	None
4	Cedrus libani	Cedar of Lebanon

[110]:
```
# Displayed the number of rows and columns
names.shape
```

[110]:
```
(589, 2)
```

[111]:
```
# Displayed a list of all the columns in the dataframe
names.columns
```

[111]:
```
Index(['Scientific Name', 'Common Name'], dtype='object')
```

[112]:
```
# Listed the data types of each column
names.dtypes
```

[112]:
```
Scientific Name    object
Common Name        object
dtype: object
```

0.2 Task 3: Further Inspect the Datasets

0.2.1 3.1 Further Inspect the Trees Dataset

Site Name

[113]:
```
# List of values in Site Name column and their counts
trees["Site Name"].value_counts()
```

[113]:
```
WATERLOW PARK (LS)                          920
Alexandra & Ainsworth Estate                289
Belsize nature reserve, Russell Nursery     278
Holly Lodge Estate                          272
LINCOLN'S INN FIELDS, GARDENS (LS)          193
                                           ...
GOLDINGTON CRESCENT                           1
ALLCROFT ROAD                                 1
WOBURN WALK, LAND BEHIND 4-18                 1
KILBURN PRIORY                                1
GOODGE PLACE                                  1
Name: Site Name, Length: 1135, dtype: int64
```

Site Name is **qualitative nominal**.

[114]:
```
#List of values for Contract Area column and its counts
trees["Contract Area"].value_counts()
```

[114]:
```
Highways              10062
Housing                7500
Parks                  4330
Education              1288
Corporate Landlord      264
Name: Contract Area, dtype: int64
```

Contract Area is **Qualitative nominal**

[115]:
```
#List of values for Scientific name and its counts
trees["Scientific Name"].value_counts()
```

[115]:
```
Platanus x hispanica                                    3340
Tilia europaea                                          1468
Acer pseudoplatanus                                      941
Betula pendula                                           765
Fraxinus excelsior                                       754
                                                         …
Vacant Tree Pit (planned: Populus tremula)                 1
Liriodendron fastigiata                                    1
Sequoia sempervirens                                       1
Sorbus x hybrida                                           1
Vacant Tree Pit (planned: Acer rubrum 'Amstrong')          1
Name: Scientific Name, Length: 543, dtype: int64
```

Scientific name is **Qualitative nominal**

[116]:
```
#List of values for Ward name and its counts
trees["Ward Name"].value_counts()
```

[116]:
```
Highgate                          2799
St Pancras and Somers Town        1832
Gospel Oak                        1541
Kilburn                           1540
Kentish Town                      1463
Haverstock                        1424
Hampstead Town                    1340
Holborn and Covent Garden         1293
Fortune Green                     1284
Cantelowes                        1231
Frognal and Fitzjohns             1229
Bloomsbury                        1008
Regent's Park                      989
Swiss Cottage                      978
Camden Town with Primrose Hill     899
West Hampstead                     853
King's Cross                       824
Belsize                            691
Name: Ward Name, dtype: int64
```

Ward Name is **Qualitative nominal**

[117]:
```
# Descriptive stats for the numeric columns

trees.describe()
```

[117]:
```
       Number Of Trees  Height In Metres  Spread In Metres  \
count     23422.000000      23006.000000      23006.000000
mean          1.100034         10.307029          5.997612
std           1.288041          6.325293          4.132379
```

```
min           0.000000          0.000000          0.000000
25%           1.000000          5.000000          3.000000
50%           1.000000          9.000000          5.000000
75%           1.000000         15.000000          8.000000
max          67.000000        127.000000         88.000000

       Diameter In Centimetres At Breast Height        Easting       Northing  \
count                              23005.000000   23444.000000   23444.000000
mean                                  32.595262  526762.518171  184085.188662
std                                   26.149994   25835.857212    9121.059755
min                                    0.000000       0.000000       0.000000
25%                                   12.000000  526583.750000  183665.000000
50%                                   27.000000  528456.500000  184690.000000
75%                                   46.000000  529369.000000  185481.000000
max                                  228.000000  531514.000000  196188.000000

          Longitude      Latitude
count  23388.000000  23388.000000
mean      -0.155145     51.545115
std        0.025058      0.013067
min       -0.261719     51.512858
25%       -0.175734     51.537472
50%       -0.148478     51.546669
75%       -0.135978     51.554032
max       -0.105858     51.650843
```

[118]:
```
# the data types
trees.dtypes
```

[118]:
```
Identifier                                      object
Number Of Trees                                float64
Site Name                                       object
Contract Area                                   object
Scientific Name                                 object
Inspection Date                         datetime64[ns]
Inspection Due Date                             object
Height In Metres                               float64
Spread In Metres                               float64
Diameter In Centimetres At Breast Height       float64
Ward Code                                       object
Ward Name                                       object
Easting                                          int64
Northing                                         int64
Longitude                                      float64
Latitude                                       float64
Location                                        object
dtype: object
```

[119]:
```
trees["Number Of Trees"].unique()
```

[119]:
```
array([ 1.,  2.,  3.,  0., nan,  5.,  6.,  7., 18.,  8., 65.,  4., 10.,
        9., 11., 50., 12., 15., 52., 40., 33., 13., 20., 67., 21., 32.,
       24., 26., 16., 25., 51.])
```

[120]:
```
trees["Height In Metres"].unique()
```

[120]:
```
array([  nan,   5. ,   4. ,  14. ,   9. ,   0. ,   2. ,   2.5,   8. ,
        13. ,  17. ,  10. ,   3. ,  19. ,   7. ,   6. ,   1.8,  15. ,
        12. ,   1.5,  16. ,  20. ,  21. ,  24. ,  25. ,   2.7,  18. ,
        11. ,  26. ,  22. ,   0.5,  16.1,  27. ,  28. ,   2.3,  22.3,
        38. ,  15.5,  29. ,  23. ,  34. ,   3.5,   2.6,   1. ,   5.5,
        22.5,   2.2,  36. ,  31. ,  96. ,  13.6, 127. ,  14.7,   4.5,
        30. ,  35. ,  32. ,   9.4,  11.8,  33. ,  40. ,   2.1,   6.5,
        41. ,  23.1,  37. ,  39. ,  12.5,   7.5,  13.7,  31.6,  23.8,
         0.2,  12.3,  15.7,   6.8,   9.3,   3.8,  24.9,  17.4,   2.8,
         3.2,  13.9,  12.4,  10.2,  10.8,  24.5,  30.4,   9.8,  11.4,
        23.4,   3.7,  11.6,   7.7,   8.3,   3.6,  17.5,  19.5])
```

[121]:
```
trees["Spread In Metres"].unique()
```

[121]:
```
array([  nan,  4.  ,  1.  ,  6.  ,  7.  ,  0.  ,  1.5 ,  5.  ,  9.  ,
       10.  ,  3.  ,  2.  , 13.  ,  2.5 , 12.  , 15.  , 14.  ,  8.  ,
       11.  ,  0.6 ,  1.8 , 17.  , 20.  , 18.  , 16.  , 22.  , 19.  ,
        5.5 , 26.  ,  1.4 , 23.  , 21.  , 28.  ,  1.2 ,  0.5 ,  3.5 ,
       24.  ,  1.3 ,  1.6 , 30.  ,  1.7 , 27.  ,  4.2 , 11.02,  0.7 ,
        6.5 ,  4.5 , 31.  , 25.  ,  0.8 ,  0.3 ,  2.2 ,  2.8 ,  1.65,
       29.  , 88.  ])
```

[122]:
```
trees["Diameter In Centimetres At Breast Height"].unique()
```

[122]:
```
array([  nan,  10. ,   6. ,  26. ,  29. ,   5. ,   0. ,   4. ,  12. ,
        59. ,  52. ,  23. ,  50. ,  63. ,  15. ,   3. ,  49. ,  42. ,
        14. ,  19. ,   9. ,  70. ,  32. ,  28. ,  34. ,  20. ,  17. ,
        27. ,  37. ,   8. ,  45. ,  18. , 119. ,   7. ,  38. ,  55. ,
        41. ,  75. ,  31. ,  25. ,  11. ,  30. ,  43. ,  68. ,  92. ,
        16. ,  35. ,  58. ,  72. ,  64. ,  13. ,  61. ,  69. ,  33. ,
        47. ,  67. , 109. , 106. ,  24. ,  51. ,  40. ,  22. ,  60. ,
        39. ,  46. ,  57. ,  21. ,  54. , 117. ,  44. ,  82. , 114. ,
        65. ,  84. , 118. ,   1. ,  89. ,  53. ,  36. ,  93. ,  88. ,
        86. , 163. ,  66. ,  81. ,  74. , 100. ,  48. ,  73. ,  95. ,
       129. ,  62. ,  85. ,  56. ,   3.5, 161. ,  71. , 105. , 102. ,
        80. ,  87. , 101. ,  76. , 113. , 108. , 160. , 132. ,  90. ,
       145. , 130. ,  79. , 228. , 110. ,  83. ,  78. , 122. , 170. ,
       115. ,   2. ,  77. , 107. ,  96. , 126. ,  91. , 104. , 158. ,
        99. ,  94. ,  16.5, 127. , 151. , 103. , 112. ,  98. ,  97. ,
       136. , 125. , 111. , 124. , 139. , 156. , 120. , 148. , 144. ,
```

```
       140. , 121. , 143. ,  17.5, 154. , 159. , 142. , 197. , 123. ,
       149. , 155. , 191. , 131. , 147. , 162. , 116. , 152. , 153. ,
       165. , 137. , 200. , 177. , 133. , 128. , 134. ,  11.5, 135. ,
       150. , 187. , 210. , 166. , 138. ,  10.5, 206. , 141. , 209. ,
       184. , 173. , 192. ,   7.5, 194. , 157. , 146. , 185. ])
```

[123]:
```
trees["Easting"].unique()
```

[123]:
```
array([527305, 529923,      0, …, 527733, 524398, 525944], dtype=int64)
```

[124]:
```
trees["Northing"].unique()
```

[124]:
```
array([185240, 184782,      0, …, 185755, 187062, 187313], dtype=int64)
```

[125]:
```
trees["Longitude"].unique()
```

[125]:
```
array([-0.16524 , -0.127681,       nan, …, -0.196884, -0.204206,
       -0.173397])
```

[126]:
```
trees["Latitude"].unique()
```

[126]:
```
array([51.551693, 51.546984,       nan, …, 51.54329 , 51.545726,
       51.531863])
```

Columns:

Number of Trees is **discrete** - Values are Int with Null's - so **NOT** a float

Height in Metres is **continuous** - Values have both int and floats

Spread in Metres is **continous** - Values have both int and floats

Diameter In Centimetres At Breast Height is **continous** - Values have more int but a few floats

Easting is **discrete** - Values are all int

Northing is **discrete** - Values are all int

Longitude is **continous** - values are float

Latitude is **continous** - values are float

0.2.2 3.2 Further Inspect the Environmental Dataset

[127]:
```
environmental["Maturity"].value_counts()
```

[127]:
```
Mature            10225
Middle aged        7779
Juvenile           4393
Not Applicable      377
Over Mature         191
Veteran              41
```

```
Name: Maturity, dtype: int64
```

Maturity is qualitative **nominal**

[128]:
```
environmental["Physiological Condition"].value_counts()
```

[128]:
```
Good              12910
Fair               9183
Poor                357
Not applicable      249
Dead                236
Excellent             8
Name: Physiological Condition, dtype: int64
```

Physiological Condition is qualitative **ordinal**

[129]:
```
environmental["Tree Set To Be Removed"].value_counts()
```

[129]:
```
No     23331
Yes       84
Name: Tree Set To Be Removed, dtype: int64
```

Trees set to be removed is qualitative **Binary**

[130]:
```
environmental["Removal Reason"].value_counts()
```

[130]:
```
Dead, dying                    30
Basal decay                    17
Trunk decay                    10
Tree defect                     5
Crown die-back                  5
Dog damage                      3
Unsuitable location             3
Newly planted tree failure      2
Coppiced stump                  1
Touching building/structure     1
Crown decay                     1
Split trunk                     1
Broken/split branch             1
Climber                         1
No defects - work required      1
Suppressed                      1
ATRD                            1
Name: Removal Reason, dtype: int64
```

Removal reason is qualitative **nominal**

3.2.b Descriptive Stats for Numeric Type Columns

[131]:
```
# Descriptive stats for all numeric columns
environmental.describe()
```

[131]:

	Capital Asset Value For Amenity Trees	Carbon Storage In Kilograms	Gross Carbon Sequestration Per Year In Kilograms	Pollution Removal Per Year In Grams
count	22982.000000	20555.000000	20555.000000	20555.000000
mean	14056.393047	467.465454	8.675057	217.740316
std	24803.806595	844.926555	8.681446	306.751920
min	0.000000	0.500000	0.000000	0.300000
25%	1035.650000	24.800000	2.200000	29.300000
50%	5443.660000	163.900000	6.100000	108.100000
75%	16781.420000	497.300000	11.700000	297.600000
max	504725.720000	6000.000000	53.800000	8223.700000

[132]:
```
# The Data Types
environmental.dtypes
```

[132]:
```
Identifier                                        object
Maturity                                          object
Physiological Condition                           object
Tree Set To Be Removed                            object
Removal Reason                                    object
Capital Asset Value For Amenity Trees            float64
Carbon Storage In Kilograms                      float64
Gross Carbon Sequestration Per Year In Kilograms    float64
Pollution Removal Per Year In Grams              float64
dtype: object
```

Find out if each numerical column values are really ints or floats

[133]:
```
environmental["Capital Asset Value For Amenity Trees"].unique()
```

[133]:
```
array([1.1507000e+02, 7.5180800e+03, 2.0419630e+04, …, 3.3664130e+04,
       3.6269450e+04, 1.4801215e+05])
```

[134]:
```
environmental["Carbon Storage In Kilograms"].unique()
```

[134]:
```
array([1.6000e+00,        nan, 4.2640e+02, …, 4.7233e+03, 3.7305e+03,
       4.8100e+02])
```

[135]:
```
environmental["Gross Carbon Sequestration Per Year In Kilograms"].unique()
```

[135]:
```
array([ 0.5,  nan,  8.8,  9.6,  1.4, 10.1,  0.8,  7.9, 24.2,  2.4,  0.9,
         3. , 14.6,  8.1,  4.1,  1.8, 18.8, 24.4, 13.4,  4.3,  6.5,  1.3,
         0.6,  2.5, 13.1,  2.1,  3.9, 15.9,  1. ,  6.6,  0.4,  3.4,  2.9,
        28.7,  0.7,  1.5, 12.5, 25.4, 11.2,  9. , 23.9,  4.4, 11.5, 11. ,
         8.4, 10.9, 10.7,  6.9,  7. ,  3.8,  3.2,  6.1,  8.6, 30.2,  3.7,
        15. , 30.4,  7.6, 20.1, 10.2,  8.3, 39.9,  0.3,  4.9, 14.3, 13.5,
        16.7,  8.9,  1.6,  4.2,  3.6,  4. ,  6.7,  0.1, 19.7, 24.6,  6.4,
         5.4,  5.9, 12.2,  7.3, 13. ,  7.1, 36.9,  9.3, 18.2, 10. ,  4.7,
         5. , 29.8, 17.8, 18.5, 17.6,  7.2,  4.8,  5.6,  5.3, 10.5, 12.9,
         9.4,  7.8,  1.1, 19.2, 37.3,  2.7, 12.8, 17. , 15.5, 27.6, 34.2,
         5.7,  5.8, 17.3, 20.4,  9.9, 15.6,  7.4, 11.8,  9.2,  2.6, 21.7,
        11.3, 10.8, 29.4, 23.1, 26.5,  1.2, 10.6, 33.6, 23.6, 11.7, 17.4,
        15.1, 16.6,  2.3, 13.3, 16.3, 18.3, 10.4, 19. ,  1.7,  3.3, 14.1,
         5.5, 12.6, 14.5,  3.5,  8.7, 36.6, 17.7,  5.1, 21.1, 20.2,  7.7,
        15.2,  7.5, 30.5, 13.8,  4.6, 39.2, 37.8, 13.2, 51. , 11.6,  6.8,
        21. ,  5.2,  8. ,  4.5, 15.3, 13.7, 18. ,  6.3,  2. ,  2.8,  9.1,
         6. , 12.4, 11.4, 18.6,  1.9, 23.5, 18.9, 26.2, 12.1, 41.7,  2.2,
        14.7, 14.4,  9.7, 23. , 22.1, 15.7, 12. , 14.8, 19.9, 29.7, 11.1,
        19.6, 12.7, 31.9, 33.2, 19.1,  6.2, 28.4, 27. , 17.5, 16.4, 27.7,
        38.4, 43.9, 35.4,  8.2, 37.1, 29.5, 22.2, 17.1, 26.3, 21.2, 20.5,
        11.9, 31.3, 14.2,  8.5, 26.8,  0.2, 16.8, 35.6, 20.8, 22.3, 19.4,
        20. , 14. , 36.1, 20.9, 15.8, 14.9,  9.5, 28.3, 16.9,  3.1, 37.6,
        34.7, 17.9, 35.3,  9.8, 16.5, 21.4, 19.3, 28.5, 29. , 33.5, 40.7,
        28.6, 13.6, 27.1, 26.7, 13.9, 21.3, 23.8, 35.1, 40.5, 32.9, 18.1,
        37. , 20.6, 35.2, 12.3, 41.8, 22.7, 21.5, 35.7, 32.4, 16.1, 31.1,
        25.3, 26. , 23.4, 30.1, 42.1, 36.5, 10.3, 44.3, 16. , 23.2, 28.8,
        21.9, 32. , 30.3, 15.4, 24.3, 26.9, 25.8, 28.2, 22.8, 42.3, 33.9,
        25.6, 38.6, 18.4, 31.2, 32.7, 28. , 30.6, 42. , 27.3, 33.8, 21.6,
        33.3, 35.9, 24. , 40.6, 42.7, 31.5, 22.5, 20.7, 33.1, 35.5, 27.5,
        24.8, 26.4, 22.6, 18.7, 32.6, 36.4, 16.2, 38.5, 22.4, 22.9, 29.2,
        25.9, 19.5, 24.5, 20.3, 27.9, 29.6, 25. , 27.4, 30.7, 25.1, 34.6,
        38. , 34.4, 36.7, 34.1, 32.5, 27.8, 32.8, 43.1, 17.2, 37.5, 39.3,
        29.1, 30.8, 34.8, 29.9, 38.3, 23.3, 21.8, 41.3, 38.7, 36.3, 30. ,
        41.5, 37.2, 24.9, 19.8, 28.9, 42.2, 39.5, 31.8, 37.4, 24.1,  0. ,
        37.7, 32.3, 36.8, 27.2, 24.7, 36.2, 39.7, 42.9, 49.5, 31.6, 38.9,
```

```
       40.9, 44.2, 32.1, 42.5, 25.7, 35.8, 38.2, 40.8, 35. , 34. , 31.7,
       40.1, 26.1, 25.2, 39. , 41.1, 33.4, 32.2, 43.3, 33.7, 26.6, 39.1,
       41.4, 43.2, 31. , 41.9, 41. , 42.4, 43. , 40.3, 40.2, 29.3, 31.4,
       22. , 42.8, 43.7, 23.7, 34.5, 41.6, 39.6, 39.8, 37.9, 43.8, 42.6,
       41.2, 43.6, 36. , 39.4, 43.5, 25.5, 38.1, 34.9, 28.1, 43.4, 53.8,
       30.9, 40.4, 34.3, 45.5, 52.7, 40. , 38.8, 44.9, 33. ])
```

[136]:
```
environmental["Pollution Removal Per Year In Grams"].unique()
```

[136]:
```
array([  5.7,   nan, 215.2, …,   8. , 399.9,  60.1])
```

Columns

Capital Asset Value For Amenity Trees is **continous** - values are float

Carbon Storage In Kilograms is **continous** - values are float

Gross Carbon Sequestration Per Year In Kilograms is **continous** - values are float

Pollution Removal Per Year In Grams is **continous** - values are float

0.2.3 3.3 Further Inspect the Common Names Dataset

3.3.a (Names) Counts of Values for String Type Columns

[137]:
```
names["Scientific Name"].value_counts()
```

[137]:
```
Cupressocyparis leylandii                              2
Larix decidua                                          2
Salix fragilis                                         2
Alnus cordata                                          2
Populus nigra                                          2
                                                      ..
Pyrus salicifolia 'Pendula'                            1
Chamaecyparis lawsoniana 'unid                         1
Platanus x hispanica Tremonia                          1
Vacant Tree Pit (planned: Gymnocladus dioicus)         1
Vacant Tree Pit (planned: Liquidambar styraciflua)     1
Name: Scientific Name, Length: 560, dtype: int64
```

Scientific name is qualitative **nominal**

[138]:
```
names["Common Name"].value_counts()
```

[138]:
```
Cherry                        12
Rowan                         10
Magnolia                      10
Vacant Tree Pit (planned: )   10
Apple - Crab                   9
                              ..
Pittosporum                    1
Birch - Purple                 1
```

```
Maple - Column Norway           1
Maple - Crimson King Norway     1
Castlewellan gold               1
Name: Common Name, Length: 431, dtype: int64
```

Common name is qualitative **nominal**

3.3.b (Names) Descriptive Stats for Numeric Type Columns There are no numeric columns.

0.3 Task 4: Identify Missing Values

0.3.1 4.1 Missing Values for the Trees Dataset

[139]:
```
# Percentage of null values
trees.isnull().mean()*100
```

[139]:
```
Identifier                                  0.000000
Number Of Trees                             0.093841
Site Name                                   0.000000
Contract Area                               0.000000
Scientific Name                             0.000000
Inspection Date                             1.710459
Inspection Due Date                         1.710459
Height In Metres                            1.868282
Spread In Metres                            1.868282
Diameter In Centimetres At Breast Height    1.872547
Ward Code                                   0.963999
Ward Name                                   0.963999
Easting                                     0.000000
Northing                                    0.000000
Longitude                                   0.238867
Latitude                                    0.238867
Location                                    0.238867
dtype: float64
```

[140]:
```
# Number of null values
trees.isnull().sum()
```

[140]:
```
Identifier                                  0
Number Of Trees                            22
Site Name                                   0
Contract Area                               0
Scientific Name                             0
Inspection Date                           401
Inspection Due Date                       401
Height In Metres                          438
Spread In Metres                          438
```

```
Diameter In Centimetres At Breast Height    439
Ward Code                                   226
Ward Name                                   226
Easting                                       0
Northing                                      0
Longitude                                    56
Latitude                                     56
Location                                     56
dtype: int64
```

[141]:
```
# Percentage of zero values
trees.isin([0]).mean()*100
```

[141]:
```
Identifier                                  0.000000
Number Of Trees                             0.396690
Site Name                                   0.000000
Contract Area                               0.000000
Scientific Name                             0.000000
Inspection Date                             0.000000
Inspection Due Date                         0.000000
Height In Metres                            0.733663
Spread In Metres                            1.181539
Diameter In Centimetres At Breast Height    1.164477
Ward Code                                   0.000000
Ward Name                                   0.000000
Easting                                     0.238867
Northing                                    0.238867
Longitude                                   0.000000
Latitude                                    0.000000
Location                                    0.000000
dtype: float64
```

[142]:
```
# Number of zero values
trees.isin([0]).sum()
```

[142]:
```
Identifier                                    0
Number Of Trees                              93
Site Name                                     0
Contract Area                                 0
Scientific Name                               0
Inspection Date                               0
Inspection Due Date                           0
Height In Metres                            172
Spread In Metres                            277
Diameter In Centimetres At Breast Height    273
Ward Code                                     0
Ward Name                                     0
```

```
Easting                                    56
Northing                                   56
Longitude                                   0
Latitude                                    0
Location                                    0
dtype: int64
```

[143]:
```
# Percentage of null and zero values
(trees.isnull().sum() + trees.isin([0]).sum())/trees.shape[0]
```

[143]:
```
Identifier                                 0.000000
Number Of Trees                            0.004905
Site Name                                  0.000000
Contract Area                              0.000000
Scientific Name                            0.000000
Inspection Date                            0.017105
Inspection Due Date                        0.017105
Height In Metres                           0.026019
Spread In Metres                           0.030498
Diameter In Centimetres At Breast Height   0.030370
Ward Code                                  0.009640
Ward Name                                  0.009640
Easting                                    0.002389
Northing                                   0.002389
Longitude                                  0.002389
Latitude                                   0.002389
Location                                   0.002389
dtype: float64
```

[144]:
```
# Number of null and zero values
(trees.isnull().sum() + trees.isin([0]).sum())
```

[144]:
```
Identifier                                   0
Number Of Trees                            115
Site Name                                    0
Contract Area                                0
Scientific Name                              0
Inspection Date                            401
Inspection Due Date                        401
Height In Metres                           610
Spread In Metres                           715
Diameter In Centimetres At Breast Height   712
Ward Code                                  226
Ward Name                                  226
Easting                                     56
Northing                                    56
Longitude                                   56
```

```
Latitude                                    56
Location                                    56
dtype: int64
```

0.3.2 4.2 Missing Values for the Environmental Dataset

[145]:
```
# Percentage of null values
# TODO
environmental.isnull().mean()*100
```

[145]:
```
Identifier                                   0.000000
Maturity                                     1.746744
Physiological Condition                      2.015802
Tree Set To Be Removed                       0.000000
Removal Reason                              99.641256
Capital Asset Value For Amenity Trees        1.849242
Carbon Storage In Kilograms                 12.214392
Gross Carbon Sequestration Per Year In Kilograms    12.214392
Pollution Removal Per Year In Grams         12.214392
dtype: float64
```

[146]:
```
# Number of null values
# TODO
environmental.isnull().sum()
```

[146]:
```
Identifier                                      0
Maturity                                      409
Physiological Condition                       472
Tree Set To Be Removed                          0
Removal Reason                              23331
Capital Asset Value For Amenity Trees         433
Carbon Storage In Kilograms                  2860
Gross Carbon Sequestration Per Year In Kilograms     2860
Pollution Removal Per Year In Grams          2860
dtype: int64
```

[147]:
```
# Percentage of zero values
# TODO
environmental.isin([0]).mean()*100
```

[147]:
```
Identifier                                   0.000000
Maturity                                     0.000000
Physiological Condition                      0.000000
Tree Set To Be Removed                       0.000000
Removal Reason                               0.000000
Capital Asset Value For Amenity Trees        1.183002
Carbon Storage In Kilograms                  0.000000
```

```
Gross Carbon Sequestration Per Year In Kilograms    0.025625
Pollution Removal Per Year In Grams                 0.000000
dtype: float64
```

[148]:
```
# Number of zero values
# TODO
environmental.isin([0]).sum()
```

[148]:
```
Identifier                                          0
Maturity                                            0
Physiological Condition                             0
Tree Set To Be Removed                              0
Removal Reason                                      0
Capital Asset Value For Amenity Trees             277
Carbon Storage In Kilograms                         0
Gross Carbon Sequestration Per Year In Kilograms    6
Pollution Removal Per Year In Grams                 0
dtype: int64
```

[149]:
```
# Percentage of null and zero values
# TODO
(environmental.isnull().sum() + environmental.isin([0]).sum())/environmental.
 ↪shape[0]
```

[149]:
```
Identifier                                          0.000000
Maturity                                            0.017467
Physiological Condition                             0.020158
Tree Set To Be Removed                              0.000000
Removal Reason                                      0.996413
Capital Asset Value For Amenity Trees               0.030322
Carbon Storage In Kilograms                         0.122144
Gross Carbon Sequestration Per Year In Kilograms    0.122400
Pollution Removal Per Year In Grams                 0.122144
dtype: float64
```

[150]:
```
# Number of null and zero values
# TODO
environmental.isnull().sum() + environmental.isin([0]).sum()
```

[150]:
```
Identifier                                            0
Maturity                                            409
Physiological Condition                             472
Tree Set To Be Removed                                0
Removal Reason                                    23331
Capital Asset Value For Amenity Trees               710
Carbon Storage In Kilograms                        2860
Gross Carbon Sequestration Per Year In Kilograms   2866
```

```
Pollution Removal Per Year In Grams                  2860
dtype: int64
```

0.3.3 4.3 Missing Values for the Common Names Dataset

[151]:
```
#Percentage of null values
names.isnull().mean()*100
```

[151]:
```
Scientific Name    0.000000
Common Name        4.074703
dtype: float64
```

[152]:
```
#Number of null values
names.isnull().sum()
```

[152]:
```
Scientific Name     0
Common Name        24
dtype: int64
```

[153]:
```
#Percentage of zero values
names.isin([0]).mean()*100
```

[153]:
```
Scientific Name    0.0
Common Name        0.0
dtype: float64
```

[154]:
```
#Number of zero values
names.isin([0]).sum()
```

[154]:
```
Scientific Name    0
Common Name        0
dtype: int64
```

[155]:
```
#Percentage of null and zero values
(names.isnull().sum() + names.isin([0]).sum())/names.shape[0]
```

[155]:
```
Scientific Name    0.000000
Common Name        0.040747
dtype: float64
```

[156]:
```
#Number of null and zero values
names.isnull().sum() + names.isin([0]).sum()
```

[156]:
```
Scientific Name     0
Common Name        24
dtype: int64
```

0.3.4 4.4 Observations

Trees data:

Columns that have the greatest null and zero values:

- Inspection Date- 401
- Inspection Due Date- 401
- Height In Metres- 610
- Spread In Metres- 715
- Diameter In Centimetres At Breast Height- 712

Columns that have a moderate null and zero value (less than 401)

- Number of Trees- 115
- Ward Code- 226
- Ward Name- 226
- Easting- 56
- Northing- 56
- Longitude- 56
- Latitude- 56
- Location- 56

Environment data:

Columns that have the greatest null and zero values:

- Removal Reason- 23331
- Carbon Storage In Kilograms- 2860
- Gross Carbon Sequestration Per Year In Kilograms- 2866
- Pollution Removal Per Year In Grams- 2860

Columns that have a moderate null and zero value (less than 2860)

- Maturity- 409
- Physiological Condition- 472
- Capital Asset Value For Amenity Trees- 710

Common names data:

Columns that have the greatest null and zero values: - Common name - 24

How can this affect the councils initiative and projects?

Public Tree Data Due to the missing data, this could affect the ammount of information displayed on the public tree data section on the council website.

Tree walks brochure Since this one is focused on a guided walk, the missing data regarding the common name, location, latitude and logitude of the trees could cause some issues since the users may not be able to find certain trees they're looking for.

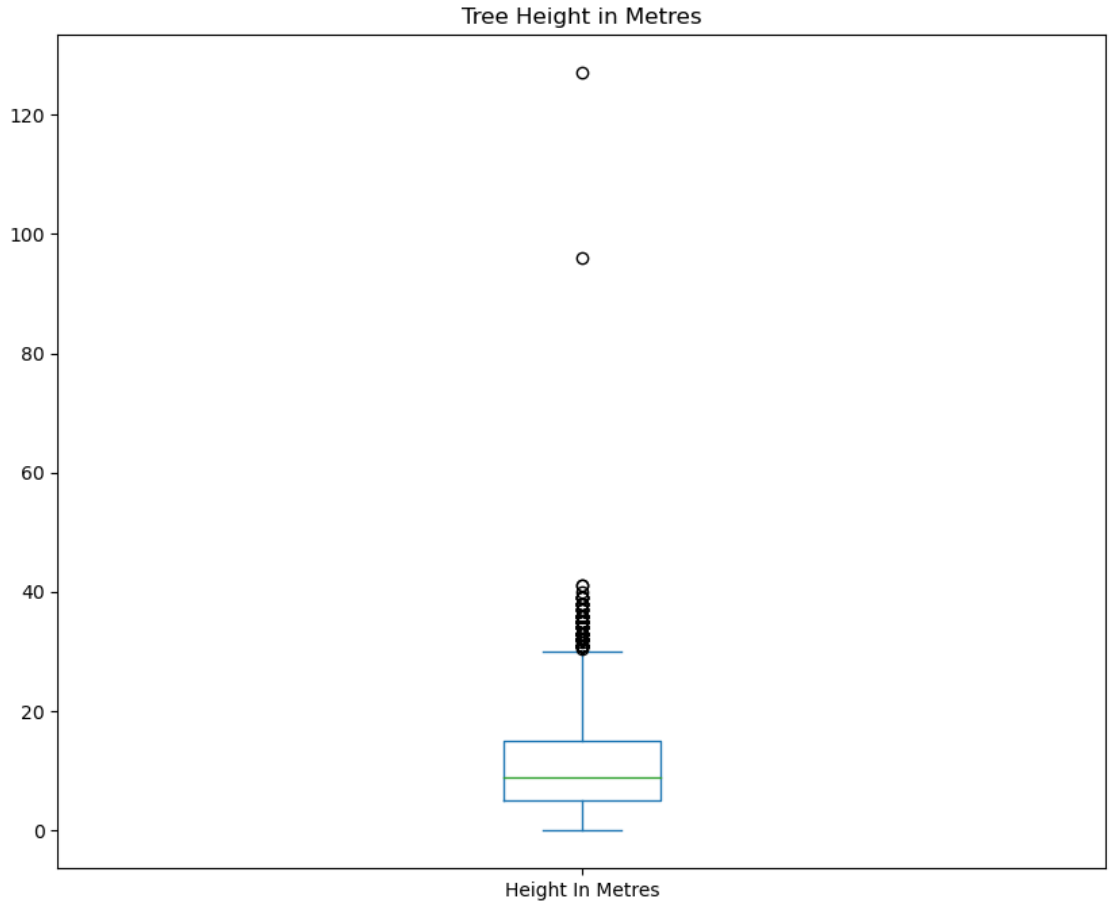
Environment Report Due to the missing data regarding the removal reason, carbon storage, pollution removal and physiological condition, this could cause bias in the environment report which could lead to inaccuracies and a issue of trust in the council.

0.4 Task 5: Identify Outliers in the Trees Dimensions

0.4.1 5.1 Outliers for Height

[157]:
```
#Used a Boxplot to find the outliers

trees["Height In Metres"].plot(kind='box', title='Tree Height in Metres',␣
↪figsize=(10,8))
```

[157]:
```
<AxesSubplot:title={'center':'Tree Height in Metres'}>
```



[158]:
```
# Selected the crazy outlier rows
mask=trees["Height In Metres"]>90
trees[mask]
```

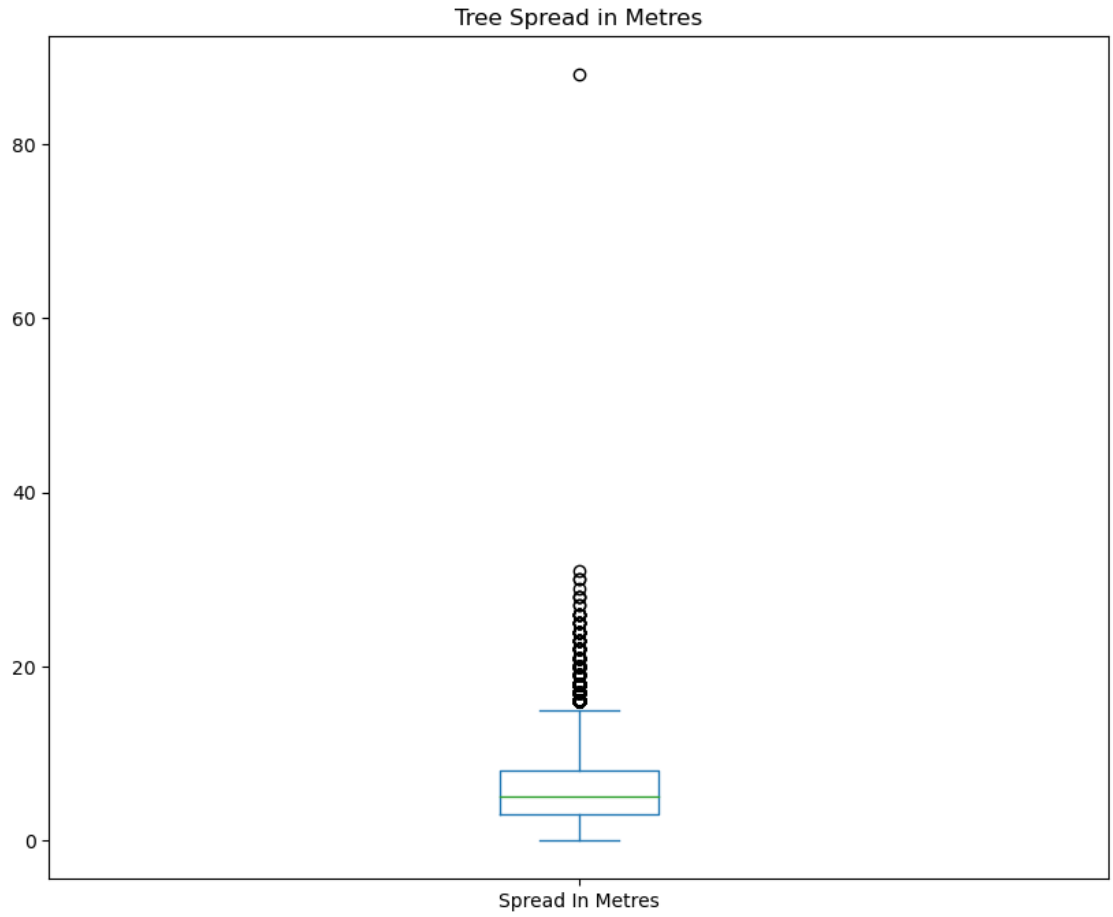
[158]:
```
     Identifier  Number Of Trees              Site Name Contract Area  \
1356   00001547              1.0    Beaumont Walk Estate       Housing
1863   00013862              1.0  Maitland Park Estate 1       Housing

           Scientific Name Inspection Date Inspection Due Date  \
1356  Robinia pseudoacacia      2017-05-23           2020/2021
1863          Prunus avium      2017-05-16           2020/2021

      Height In Metres  Spread In Metres  \
1356              96.0              10.0
1863             127.0               9.0

      Diameter In Centimetres At Breast Height  Ward Code   Ward Name  \
1356                                      63.0  E05000136  Haverstock
```

```
1863                                    34.0  E05000136  Haverstock

      Easting  Northing  Longitude   Latitude                Location
1356   527847    184391  -0.157739  51.543941  (51.543941, -0.157739)
1863   527987    184901  -0.155534  51.548497  (51.548497, -0.155534)
```

0.4.2 5.2 Outliers for Spread

[159]:
```
# Used a boxplot to find the outliers
trees["Spread In Metres"].plot(kind='box', title='Tree Spread in Metres',␣
↪figsize=(10,8))
```

[159]:
```
<AxesSubplot:title={'center':'Tree Spread in Metres'}>
```



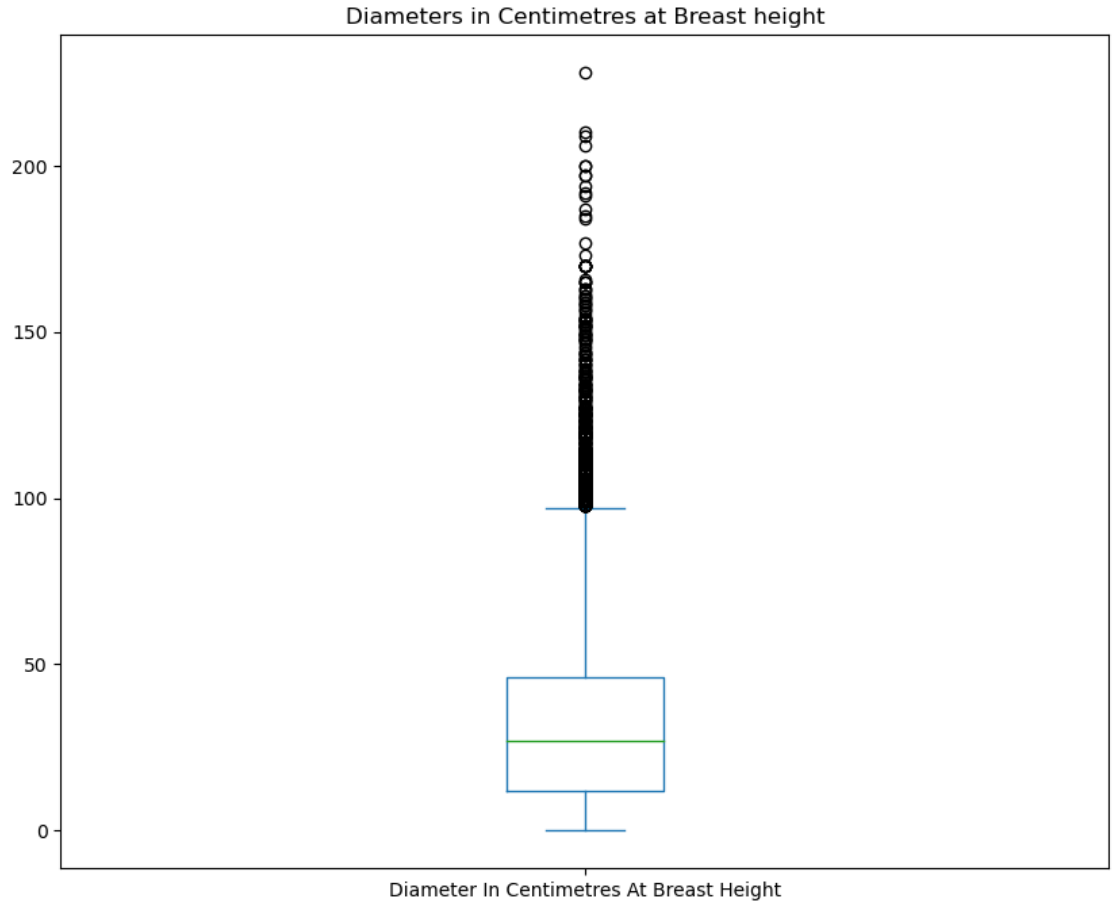
[160]:
```
# Selected the crazy outlier rows (if any)
mask=trees["Spread In Metres"]>80
trees[mask]
```

[160]:
```
       Identifier  Number Of Trees            Site Name Contract Area  \
18567   00045515              1.0  Broadfield Estate 1       Housing

      Scientific Name Inspection Date Inspection Due Date  Height In Metres  \
18567   Quercus robur      2018-04-26           2021/2022               8.0

       Spread In Metres  Diameter In Centimetres At Breast Height  Ward Code  \
18567              88.0                                      17.0  E05000144

           Ward Name  Easting  Northing  Longitude   Latitude  \
18567  Swiss Cottage   525993    184693  -0.184348  51.547074

                     Location
18567  (51.547074, -0.184348)
```

0.4.3 5.3 Outliers for Diameter

[161]:
```
# Used a boxplot to find the outliers
trees["Diameter In Centimetres At Breast Height"].plot(kind='box',␣
 ↪title='Diameters in Centimetres at Breast height', figsize=(10,8))
```

[161]:
```
<AxesSubplot:title={'center':'Diameters in Centimetres at Breast height'}>
```



[162]:
```
# Selected the outlier rows (if any)
mask=trees["Diameter In Centimetres At Breast Height"]>170
trees[mask]
```

[162]:

	Identifier	Number Of Trees	Site Name
1157	00004100	1.0	LONGFORD STREET, CLARENCE GDNS (LS)
5229	00002735	1.0	BRUNSWICK SQUARE, GARDENS (LS)
5859	00017570	1.0	RUSSELL SQUARE, GARDENS (LS)
8179	00012846	1.0	LINCOLN'S INN FIELDS, GARDENS (LS)
8737	00019491	1.0	PANCRAS ROAD, ST. PANCRAS GDNS (LS)
8870	00017656	1.0	RUSSELL SQUARE, GARDENS (LS)
11665	00020242	1.0	TAVISTOCK SQUARE, GARDENS (LS)
11860	00012885	1.0	LINCOLN'S INN FIELDS, GARDENS (LS)
14178	00012891	1.0	LINCOLN'S INN FIELDS, GARDENS (LS)
15853	00012939	1.0	LINCOLN'S INN FIELDS, GARDENS (LS)
16200	00012833	1.0	LINCOLN'S INN FIELDS, GARDENS (LS)
17816	00021404	1.0	WATERLOW PARK (LS)
18081	00012870	1.0	LINCOLN'S INN FIELDS, GARDENS (LS)

```
18930   00018775              1.0  ST. GEORGE'S GARDENS, HEATHCOTE ST (LS)
20426   00012873              1.0       LINCOLN'S INN FIELDS, GARDENS (LS)
22781   00012923              1.0       LINCOLN'S INN FIELDS, GARDENS (LS)

      Contract Area       Scientific Name Inspection Date Inspection Due Date  \
1157          Parks  Platanus x hispanica      2018-03-29           2020/2021
5229          Parks  Platanus x hispanica      2018-03-23           2020/2021
5859          Parks  Platanus x hispanica      2018-03-14           2020/2021
8179          Parks  Platanus x hispanica      2018-04-24           2021/2022
8737          Parks  Platanus x hispanica      2017-03-14           2019/2020
8870          Parks  Platanus x hispanica      2018-03-20           2020/2021
11665         Parks  Platanus x hispanica      2018-06-04           2021/2022
11860         Parks  Platanus x hispanica      2018-04-20           2021/2022
14178         Parks  Platanus x hispanica      2018-04-20           2021/2022
15853         Parks  Platanus x hispanica      2018-04-19           2021/2022
16200         Parks  Platanus x hispanica      2018-04-25           2021/2022
17816         Parks  Platanus x hispanica      2019-06-19           2022/2023
18081         Parks  Platanus x hispanica      2018-04-20           2021/2022
18930         Parks  Platanus x hispanica      2018-03-20           2020/2021
20426         Parks  Platanus x hispanica      2018-04-20           2021/2022
22781         Parks  Platanus x hispanica      2018-04-19           2021/2022

       Height In Metres  Spread In Metres  \
1157               21.0              20.0
5229               28.0              30.0
5859               38.0              16.0
8179               28.0              25.0
8737               24.0              25.0
8870               36.0              21.0
11665              28.0              24.0
11860              24.0              24.0
14178              30.0              20.0
15853              23.0              20.0
16200              29.0              22.0
17816              30.0              25.0
18081              28.0              18.0
18930              35.0              24.0
20426              29.0              26.0
22781              27.0              22.0

       Diameter In Centimetres At Breast Height  Ward Code  \
1157                                      228.0  E05000142
5229                                      197.0  E05000141
5859                                      191.0  E05000129
8179                                      200.0  E05000138
8737                                      200.0  E05000143
8870                                      177.0  E05000129
```

```
11665                                    187.0  E05000129
11860                                    210.0  E05000138
14178                                    206.0  E05000138
15853                                    209.0  E05000138
16200                                    184.0  E05000138
17816                                    173.0  E05000137
18081                                    192.0  E05000138
18930                                    194.0  E05000141
20426                                    197.0  E05000138
22781                                    185.0  E05000138

                      Ward Name  Easting  Northing  Longitude   Latitude  \
1157                Regent's Park   528931    182624  -0.142766  51.527816
5229                 King's Cross   530387    182261  -0.121914  51.524218
5859                   Bloomsbury   530025    181952  -0.127255  51.521529
8179    Holborn and Covent Garden   530834    181409  -0.115802  51.516467
8737   St Pancras and Somers Town   529789    183478  -0.130079  51.535299
8870                   Bloomsbury   530122    181955  -0.125851  51.521536
11665                  Bloomsbury   529873    182306  -0.129313  51.524743
11860   Holborn and Covent Garden   530820    181358  -0.116011  51.516010
14178   Holborn and Covent Garden   530783    181341  -0.116553  51.515861
15853   Holborn and Covent Garden   530705    181373  -0.117672  51.516165
16200   Holborn and Covent Garden   530816    181440  -0.116050  51.516750
17816                     Highgate   528815    187160  -0.142769  51.568608
18081   Holborn and Covent Garden   530873    181365  -0.115257  51.516056
18930                King's Cross   530485    182497  -0.120423  51.526317
20426   Holborn and Covent Garden   530853    181378  -0.115534  51.516179
22781   Holborn and Covent Garden   530769    181399  -0.116734  51.516390

                     Location
1157   (51.527816, -0.142766)
5229   (51.524218, -0.121914)
5859   (51.521529, -0.127255)
8179   (51.516467, -0.115802)
8737   (51.535299, -0.130079)
8870   (51.521536, -0.125851)
11665  (51.524743, -0.129313)
11860   (51.51601, -0.116011)
14178  (51.515861, -0.116553)
15853  (51.516165, -0.117672)
16200    (51.51675, -0.11605)
17816  (51.568608, -0.142769)
18081  (51.516056, -0.115257)
18930  (51.526317, -0.120423)
20426  (51.516179, -0.115534)
22781   (51.51639, -0.116734)
```

0.4.4 5.4 Observations

The outliers in all three columns (Height in Metres, Spread in Metres and Diameter In Centimetres At Breast Height) were easy to identify. This was because I noticed that most of the data points above the maximum were tightly connected to eachother, so the data points that were further away and segregated must have been the outlier since it wasn't linked to the previous data points.

The column that had the most outliers was the Diameter in centimetres at breast height.However, the outliers for the Diameter in cm at breast height is difficult to justify whether or not it is a incorrect input since all the outliers seem to be linked closely. Hence, I chose 170 as my threshold value for the outliers in this case.

The Height and Spread columns seemed to show the fewest outliers.

The Height columns outlier must have been inputted correctly, since there are no trees above the height of 100 metres in Camden. This matter could be possible for the Spread column too since the outlier is segregated from the rest of the circles, which could indicate a incorrect value too.

0.5 Task 6: Identify Duplicates in the Trees Dataset

0.5.1 6.1 Find Duplicate Rows

In our dataset the `Identifier` column should be unique. Find out if it is! We've already used a function that can count how many times each value in a column exists. Use is to see if we have duplicates in the trees Dataframe.

[163]:
```
# Found out if we have any duplicates
trees["Identifier"].duplicated().sum()
```

[163]:
```
6
```

[164]:
```
identifier_counts = trees["Identifier"].value_counts()
identifier_counts[identifier_counts>1]
```

[164]:
```
00000999    2
00060087    2
00022744    2
00032549    2
00022674    2
00060088    2
Name: Identifier, dtype: int64
```

Now see if you can select the rows from trees DataFrame that are duplicates. You will need to use the output from the cell above and use it to filter the trees dataframe.

[165]:
```
# Selected the rows that are duplicated
trees[trees["Identifier"].duplicated()]
```

[165]:
```
      Identifier  Number Of Trees         Site Name Contract Area  \
6111    00032549              1.0    NARCISSUS ROAD      Highways
9186    00022744              1.0          YORK WAY      Highways
```

```
10972   00060088              1.0  FREDERICK STREET      Highways
13098   00000999              1.0       ALMA STREET      Highways
13628   00022674              1.0     WOODSOME ROAD      Highways
15653   00060087              1.0     ARGYLE SQUARE      Highways

                                          Scientific Name Inspection Date  \
6111                           Betula albosinensis Fasc.      2018-09-19
9186                                 Ailanthus altissima      2019-10-30
10972  Vacant Tree Pit (planned: Acer campestre eco s…      2019-11-09
13098                                Sorbus hupehensis      2017-07-25
13628                                           Sorbus      2017-10-07
15653  Vacant Tree Pit (planned: Acer campestre eco s…      2019-11-09

      Inspection Due Date  Height In Metres  Spread In Metres  \
6111            2021/2022               9.0               6.0
9186            2022/2023               7.5               3.0
10972           2022/2023               NaN               NaN
13098           2020/2021               5.0               4.0
13628           2020/2021               7.0               6.0
15653           2022/2023               NaN               NaN

       Diameter In Centimetres At Breast Height  Ward Code       Ward Name  \
6111                                       19.0  E05000145  West Hampstead
9186                                       18.0  E05000131      Cantelowes
10972                                       NaN  E05000141    King's Cross
13098                                      18.0  E05000139    Kentish Town
13628                                      28.0  E05000137        Highgate
15653                                       NaN  E05000141    King's Cross

       Easting  Northing  Longitude   Latitude                Location
6111    525185    185127  -0.195840  51.551159   (51.551159, -0.19584)
9186    529983    184724  -0.126837  51.546447  (51.546447, -0.126837)
10972   530770    182696  -0.116241  51.528046  (51.528046, -0.116241)
13098   528834    184856  -0.143337  51.547904  (51.547904, -0.143337)
13628   528515    186109  -0.147481  51.559237  (51.559237, -0.147481)
15653   530342    182839  -0.122352  51.529428  (51.529428, -0.122352)
```

0.5.2 6.2 Observations

There were 6 duplicates in the Identifier column in the Tree data set. Luckily, it was only duplicated once. The identifiers that had the duplicates were: 00000999
00060087
00022744
00032549
00022674
00060088

These duplicates would have to be removed from the data set before they are used in the council's

projects or during analysis since this could cause inaccuracies in the data and subsequent analysis

0.6 Task 7: Identify Geolocation Issues

The geographic coordinates (Easting and Northing) can be used to plot the trees on a map. We can use this approach to see if there are any unusual tree locations!

0.6.1 7.1 Remove Trees with Missing Geo-coordinates

[166]:
```
# Made a copy of the trees
geotrees = trees.copy()
```

[167]:
```
# Removed null Eastings

mask = geotrees["Easting"].notnull()
geotrees[mask]
```

[167]:
```
      Identifier  Number Of Trees                          Site Name  \
0       00060053              1.0           Russell Nurseries Estate
1       00057855              1.0                  BRECKNOCK JMI (E)
2       00059953              1.0         Estate 51 Ravenshaw Street
3       00059915              1.0                   ROSARY RC JMI (E)
4       00010762              1.0                 Holly Lodge Estate
…          …                  …                                  …
23439   00057455              1.0  KILBURN GRANGE, MESSINA AVE (LS)
23440   00015494              1.0                   OSSULSTON STREET
23441   00001693              1.0                     BELMONT STREET
23442   00020342              1.0               Templar House Estate
23443   00013369              1.0                     LYNDHURST ROAD

      Contract Area          Scientific Name Inspection Date  \
0           Housing          Vacant Tree Pit             NaT
1         Education          Vacant Tree Pit      2019-07-17
2           Housing             Ficus carica             NaT
3         Education      Betula jacquemontii             NaT
4           Housing     Ilex x altaclarensis      2017-06-14
…             …                        …              …
23439         Parks              Sorbus aria      2017-03-28
23440      Highways  Sorbus aria 'Majestica'      2019-10-28
23441      Highways     Platanus x hispanica      2017-07-31
23442       Housing           Tilia europaea      2018-11-05
23443      Highways           Tilia europaea      2018-06-08

      Inspection Due Date  Height In Metres  Spread In Metres  \
0                     NaN               NaN               NaN
1               2022/2023               NaN               NaN
2                     NaN               5.0               4.0
3                     NaN               4.0               1.0
```

```
4               2020/2021              14.0               6.0
…                   …                  …                  …
23439           2019/2020               2.0               1.0
23440           2022/2023              12.0               6.0
23441           2020/2021              18.0              10.0
23442           2021/2022              20.0               8.0
23443           2021/2022              15.0               6.0

       Diameter In Centimetres At Breast Height  Ward Code  \
0                                           NaN  E05000135
1                                           NaN  E05000131
2                                          10.0        NaN
3                                           6.0  E05000135
4                                          26.0  E05000137
…                                           …        …
23439                                       6.0  E05000140
23440                                      48.0  E05000143
23441                                      57.0  E05000136
23442                                      40.0  E05000132
23443                                      77.0  E05000135

                        Ward Name  Easting  Northing  Longitude   Latitude  \
0                  Hampstead Town   527305    185240  -0.165240  51.551693
1                      Cantelowes   529923    184782  -0.127681  51.546984
2                             NaN        0         0        NaN        NaN
3                  Hampstead Town   527249    185261  -0.166051  51.551901
4                        Highgate   528414    186770  -0.148704  51.565198
…                             …       …         …         …         …
23439                     Kilburn   525130    184418  -0.196884  51.544796
23440  St Pancras and Somers Town   529758    183095  -0.130667  51.531863
23441                   Haverstock   528302    184457  -0.151163  51.544432
23442                Fortune Green   524615    184714  -0.204206  51.547573
23443                Hampstead Town   526738    185304  -0.173397  51.552397

                     Location
0       (51.551693, -0.16524)
1      (51.546984, -0.127681)
2                         NaN
3      (51.551901, -0.166051)
4      (51.565198, -0.148704)
…                         …
23439  (51.544796, -0.196884)
23440  (51.531863, -0.130667)
23441  (51.544432, -0.151163)
23442  (51.547573, -0.204206)
23443  (51.552397, -0.173397)
```

[23444 rows x 17 columns]

[168]:
```
# Removed 0 Eastings.
mask = geotrees["Easting"] !=0
geotrees[mask]
```

[168]:

	Identifier	Number Of Trees	Site Name
0	00060053	1.0	Russell Nurseries Estate
1	00057855	1.0	BRECKNOCK JMI (E)
3	00059915	1.0	ROSARY RC JMI (E)
4	00010762	1.0	Holly Lodge Estate
5	00007523	1.0	Westcroft Estate 1
…	…	…	…
23439	00057455	1.0	KILBURN GRANGE, MESSINA AVE (LS)
23440	00015494	1.0	OSSULSTON STREET
23441	00001693	1.0	BELMONT STREET
23442	00020342	1.0	Templar House Estate
23443	00013369	1.0	LYNDHURST ROAD

	Contract Area	Scientific Name	Inspection Date
0	Housing	Vacant Tree Pit	NaT
1	Education	Vacant Tree Pit	2019-07-17
3	Education	Betula jacquemontii	NaT
4	Housing	Ilex x altaclarensis	2017-06-14
5	Housing	Betula pendula	2018-08-06
…	…	…	…
23439	Parks	Sorbus aria	2017-03-28
23440	Highways	Sorbus aria 'Majestica'	2019-10-28
23441	Highways	Platanus x hispanica	2017-07-31
23442	Housing	Tilia europaea	2018-11-05
23443	Highways	Tilia europaea	2018-06-08

	Inspection Due Date	Height In Metres	Spread In Metres
0	NaN	NaN	NaN
1	2022/2023	NaN	NaN
3	NaN	4.0	1.0
4	2020/2021	14.0	6.0
5	2021/2022	9.0	7.0
…	…	…	…
23439	2019/2020	2.0	1.0
23440	2022/2023	12.0	6.0
23441	2020/2021	18.0	10.0
23442	2021/2022	20.0	8.0
23443	2021/2022	15.0	6.0

	Diameter In Centimetres At Breast Height	Ward Code
0	NaN	E05000135

```
1                                          NaN  E05000131
3                                          6.0  E05000135
4                                         26.0  E05000137
5                                         29.0        NaN
…                                          …          …
23439                                      6.0  E05000140
23440                                     48.0  E05000143
23441                                     57.0  E05000136
23442                                     40.0  E05000132
23443                                     77.0  E05000135

                         Ward Name  Easting  Northing  Longitude   Latitude  \
0                  Hampstead Town   527305    185240  -0.165240  51.551693
1                      Cantelowes   529923    184782  -0.127681  51.546984
3                  Hampstead Town   527249    185261  -0.166051  51.551901
4                       Highgate   528414    186770  -0.148704  51.565198
5                            NaN   524253    185982  -0.208975  51.559049
…                            …       …         …         …         …
23439                    Kilburn   525130    184418  -0.196884  51.544796
23440  St Pancras and Somers Town   529758    183095  -0.130667  51.531863
23441                  Haverstock   528302    184457  -0.151163  51.544432
23442               Fortune Green   524615    184714  -0.204206  51.547573
23443               Hampstead Town   526738    185304  -0.173397  51.552397

                     Location
0       (51.551693, -0.16524)
1      (51.546984, -0.127681)
3      (51.551901, -0.166051)
4      (51.565198, -0.148704)
5      (51.559049, -0.208975)
…                         …
23439  (51.544796, -0.196884)
23440  (51.531863, -0.130667)
23441  (51.544432, -0.151163)
23442  (51.547573, -0.204206)
23443  (51.552397, -0.173397)

[23388 rows x 17 columns]
```

[169]:
```
# Removed null Northings
mask = geotrees["Northing"].notnull()
geotrees[mask]
```

[169]:
```
      Identifier  Number Of Trees                          Site Name  \
0       00060053              1.0          Russell Nurseries Estate
1       00057855              1.0                 BRECKNOCK JMI (E)
2       00059953              1.0        Estate 51 Ravenshaw Street
```

```
3       00059915              1.0                  ROSARY RC JMI (E)
4       00010762              1.0                Holly Lodge Estate
…          …                  …                                …
23439   00057455              1.0  KILBURN GRANGE, MESSINA AVE (LS)
23440   00015494              1.0                   OSSULSTON STREET
23441   00001693              1.0                     BELMONT STREET
23442   00020342              1.0              Templar House Estate
23443   00013369              1.0                    LYNDHURST ROAD

      Contract Area          Scientific Name Inspection Date  \
0           Housing          Vacant Tree Pit             NaT
1         Education          Vacant Tree Pit      2019-07-17
2           Housing             Ficus carica             NaT
3         Education      Betula jacquemontii             NaT
4           Housing     Ilex x altaclarensis      2017-06-14
…              …                        …              …
23439         Parks              Sorbus aria      2017-03-28
23440      Highways  Sorbus aria 'Majestica'      2019-10-28
23441      Highways     Platanus x hispanica      2017-07-31
23442       Housing           Tilia europaea      2018-11-05
23443      Highways           Tilia europaea      2018-06-08

      Inspection Due Date  Height In Metres  Spread In Metres  \
0                     NaN               NaN               NaN
1               2022/2023               NaN               NaN
2                     NaN               5.0               4.0
3                     NaN               4.0               1.0
4               2020/2021              14.0               6.0
…                     …                 …                 …
23439           2019/2020               2.0               1.0
23440           2022/2023              12.0               6.0
23441           2020/2021              18.0              10.0
23442           2021/2022              20.0               8.0
23443           2021/2022              15.0               6.0

       Diameter In Centimetres At Breast Height  Ward Code  \
0                                           NaN  E05000135
1                                           NaN  E05000131
2                                          10.0        NaN
3                                           6.0  E05000135
4                                          26.0  E05000137
…                                             …          …
23439                                       6.0  E05000140
23440                                      48.0  E05000143
23441                                      57.0  E05000136
23442                                      40.0  E05000132
23443                                      77.0  E05000135
```

```
                        Ward Name  Easting  Northing  Longitude   Latitude  \
0                  Hampstead Town   527305    185240  -0.165240  51.551693
1                      Cantelowes   529923    184782  -0.127681  51.546984
2                             NaN        0         0        NaN        NaN
3                  Hampstead Town   527249    185261  -0.166051  51.551901
4                        Highgate   528414    186770  -0.148704  51.565198
…                              …      …       …         …          …
23439                      Kilburn   525130    184418  -0.196884  51.544796
23440  St Pancras and Somers Town   529758    183095  -0.130667  51.531863
23441                   Haverstock   528302    184457  -0.151163  51.544432
23442                Fortune Green   524615    184714  -0.204206  51.547573
23443                Hampstead Town   526738    185304  -0.173397  51.552397

                     Location
0       (51.551693, -0.16524)
1      (51.546984, -0.127681)
2                         NaN
3      (51.551901, -0.166051)
4      (51.565198, -0.148704)
…                         …
23439  (51.544796, -0.196884)
23440  (51.531863, -0.130667)
23441  (51.544432, -0.151163)
23442  (51.547573, -0.204206)
23443  (51.552397, -0.173397)

[23444 rows x 17 columns]
```

[170]:
```
# Removed 0 Northings.
mask = geotrees["Northing"] !=0
geotrees[mask]
```

[170]:
```
      Identifier  Number Of Trees                          Site Name  \
0       00060053              1.0          Russell Nurseries Estate
1       00057855              1.0                 BRECKNOCK JMI (E)
3       00059915              1.0                 ROSARY RC JMI (E)
4       00010762              1.0                Holly Lodge Estate
5       00007523              1.0                Westcroft Estate 1
…          …                  …                                 …
23439   00057455              1.0  KILBURN GRANGE, MESSINA AVE (LS)
23440   00015494              1.0                  OSSULSTON STREET
23441   00001693              1.0                    BELMONT STREET
23442   00020342              1.0              Templar House Estate
23443   00013369              1.0                    LYNDHURST ROAD

      Contract Area          Scientific Name Inspection Date  \
```

```
0           Housing          Vacant Tree Pit             NaT
1         Education          Vacant Tree Pit      2019-07-17
3         Education      Betula jacquemontii             NaT
4           Housing     Ilex x altaclarensis      2017-06-14
5           Housing           Betula pendula      2018-08-06
…             …                   …               …
23439         Parks              Sorbus aria      2017-03-28
23440      Highways  Sorbus aria 'Majestica'      2019-10-28
23441      Highways     Platanus x hispanica      2017-07-31
23442       Housing           Tilia europaea      2018-11-05
23443      Highways           Tilia europaea      2018-06-08

      Inspection Due Date  Height In Metres  Spread In Metres  \
0                     NaN               NaN               NaN
1               2022/2023               NaN               NaN
3                     NaN               4.0               1.0
4               2020/2021              14.0               6.0
5               2021/2022               9.0               7.0
…                     …                 …                 …
23439           2019/2020               2.0               1.0
23440           2022/2023              12.0               6.0
23441           2020/2021              18.0              10.0
23442           2021/2022              20.0               8.0
23443           2021/2022              15.0               6.0

       Diameter In Centimetres At Breast Height  Ward Code  \
0                                           NaN  E05000135
1                                           NaN  E05000131
3                                           6.0  E05000135
4                                          26.0  E05000137
5                                          29.0        NaN
…                                             …         …
23439                                       6.0  E05000140
23440                                      48.0  E05000143
23441                                      57.0  E05000136
23442                                      40.0  E05000132
23443                                      77.0  E05000135

                        Ward Name  Easting  Northing  Longitude   Latitude  \
0                  Hampstead Town   527305    185240  -0.165240  51.551693
1                      Cantelowes   529923    184782  -0.127681  51.546984
3                  Hampstead Town   527249    185261  -0.166051  51.551901
4                       Highgate   528414    186770  -0.148704  51.565198
5                            NaN   524253    185982  -0.208975  51.559049
…                              …        …         …          …          …
23439                     Kilburn   525130    184418  -0.196884  51.544796
23440  St Pancras and Somers Town   529758    183095  -0.130667  51.531863
```

```
23441                 Haverstock   528302    184457  -0.151163  51.544432
23442              Fortune Green   524615    184714  -0.204206  51.547573
23443            Hampstead Town    526738    185304  -0.173397  51.552397

                      Location
0       (51.551693, -0.16524)
1      (51.546984, -0.127681)
3      (51.551901, -0.166051)
4      (51.565198, -0.148704)
5      (51.559049, -0.208975)
…                         …
23439  (51.544796, -0.196884)
23440  (51.531863, -0.130667)
23441  (51.544432, -0.151163)
23442  (51.547573, -0.204206)
23443  (51.552397, -0.173397)

[23388 rows x 17 columns]
```

[171]:
```
# Confirmed how many rows we have
geotrees[mask].shape
```

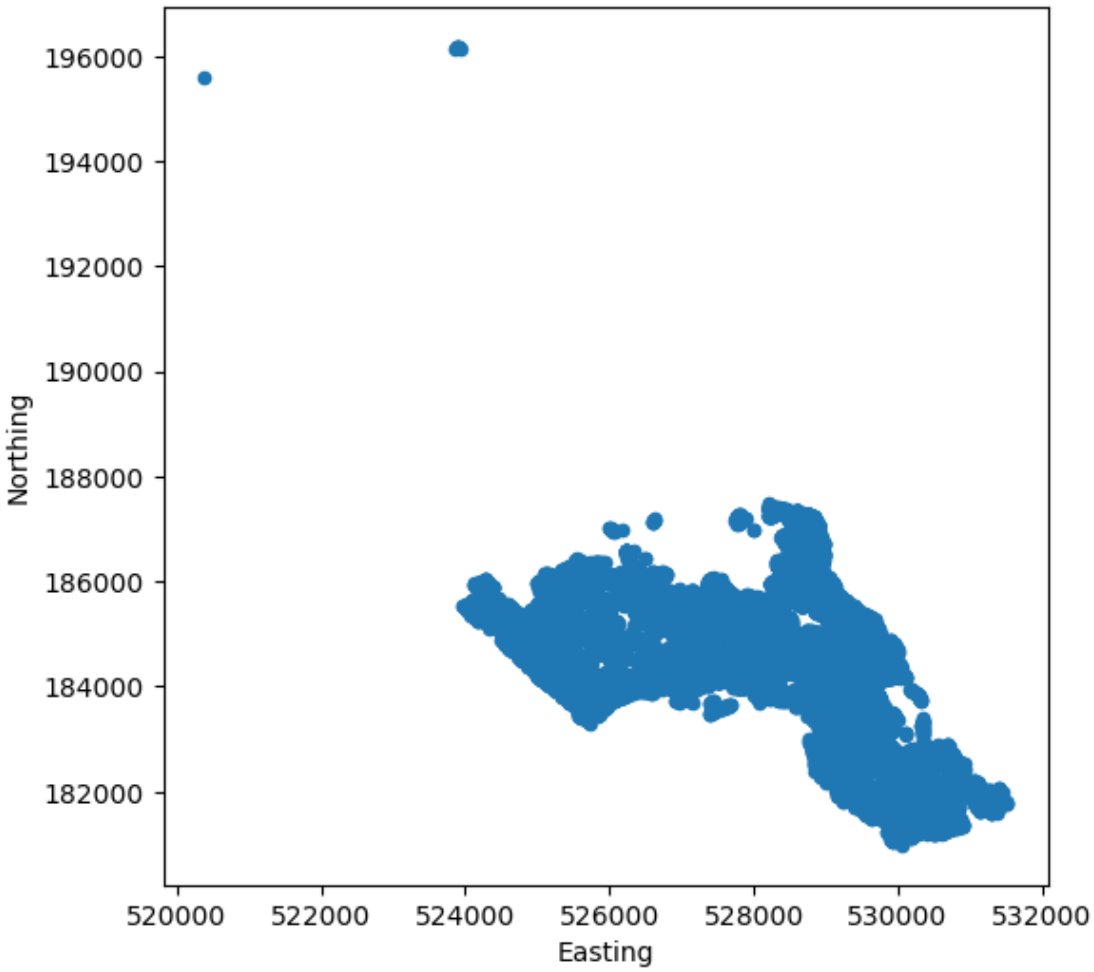
[171]:
```
(23388, 17)
```

There were 56 rows in both Easting and Northing that had 0's in the values. Both had 0 rows that contained nulls. Before the null and zero values were removed, the Count for both Easting and Northing was 23,444 but **after it was removed**, there are now **23,338** rows.

0.6.2 7.2 Plot Map of Trees

[172]:
```
# Ploted the trees on a map
geotrees[mask].plot.scatter(x='Easting', y="Northing", figsize=(6,6))
```

[172]:
```
<AxesSubplot:xlabel='Easting', ylabel='Northing'>
```



0.6.3 7.3 Find Trees Outside Camden

From the scatter plot, you should be able to determine how to select the rows from the trees data set containing the offending trees (using the Easting and Northing values)

Select the rows containing trees outside of Camden. Use the filter technique again.

[173]:
```
# Selected the outlier rows
mask = geotrees["Northing"] >194000
geotrees[mask]
```

[173]:

	Identifier	Number Of Trees	Site Name	Contract Area	\
78	00044991	1.0	Bells Hill Estate	Housing	
298	00045000	1.0	Estate 167 Furzehill Road	Housing	
660	00044992	1.0	Bells Hill Estate	Housing	
1526	00044995	1.0	Bells Hill Estate	Housing	
5392	00044990	1.0	Bells Hill Estate	Housing	

```
18069   00044993              1.0          Bells Hill Estate       Housing
18078   00044601              1.0          Bells Hill Estate       Housing
19532   00044988              1.0          Bells Hill Estate       Housing

              Scientific Name Inspection Date Inspection Due Date  \
78         Fraxinus excelsior      2017-04-28           2020/2021
298          Pinus sylvestris      2017-04-28           2020/2021
660        Crataegus monogyna      2017-04-28           2020/2021
1526       Fraxinus excelsior      2017-04-28           2020/2021
5392   Aesculus hippocastanum      2017-04-28           2020/2021
18069           Tilia cordata      2017-04-28           2020/2021
18078           Tilia cordata      2017-04-28           2020/2021
19532           Tilia cordata      2017-04-28           2020/2021

       Height In Metres  Spread In Metres  \
78                 20.0              10.0
298                10.0               3.0
660                 4.0               4.0
1526               18.0              12.0
5392               22.0              12.0
18069              23.0              14.0
18078              21.0              12.0
19532              21.0              12.0

       Diameter In Centimetres At Breast Height Ward Code Ward Name  Easting  \
78                                        49.0       NaN       NaN   523883
298                                       41.0       NaN       NaN   520367
660                                        7.0       NaN       NaN   523875
1526                                      54.0       NaN       NaN   523936
5392                                      67.0       NaN       NaN   523889
18069                                     89.0       NaN       NaN   523867
18078                                     56.0       NaN       NaN   523905
19532                                     65.0       NaN       NaN   523909

       Northing  Longitude   Latitude                Location
78       196179  -0.210713  51.650765  (51.650765, -0.210713)
298      195595  -0.261719  51.646283  (51.646283, -0.261719)
660      196170  -0.210830  51.650690    (51.65069, -0.21083)
1526     196127  -0.209969  51.650287  (51.650287, -0.209969)
5392     196188  -0.210618  51.650843  (51.650843, -0.210618)
18069    196159  -0.210957  51.650590   (51.65059, -0.210957)
18078    196174  -0.210394  51.650719  (51.650719, -0.210394)
19532    196169  -0.210340  51.650676   (51.650676, -0.21034)
```

[174]:
```
# Confirmed how many rows we have
geotrees[mask].shape
```

[174]: (8, 17)

0.6.4 7.4 Observations

For the outliers in this scatter plot, there seems to be 8 trees that are outside Camden. It was easy to identify the outliers in this case as there are two distinct dots that are segregated from the rest of the "map" of camden, which shows they aren't connected. I found the outlier by using the Northing column and by masking values that were greater than 194000 as shown on the scatter plot.

0.7 Task 8: Identify Unmatched Data

We have multiple datasets that will need to be joined together to produce the analyses required by the Camden Parks and Open Spaces team. The data will need to be joined in the following way:

- Use the `Identifier` column in the trees dataset to match to the `Identifier` column in the environmental data set (so we can bring in the environmental data for each tree)
- Use the `Scientific Name` column in the trees dataset to match to the `Scientific Name` column in the common names data set (so we can look up the `Common Name`)

There may be mismatches in the data. Of particular concern we want to check

- That every tree in the trees dataset has matching environmental data in the environmental data set
- That every environmental row in the environmental dataset has matching tree data in the tree data set
- That every scientific name in the trees dataset has a matching common name in the common names data set

0.7.1 8.1 Find Trees that Don't have Matching Environmental Data

[175]:
```
# Found trees that don't have matching environmental data

mask = ~trees["Identifier"].isin(environmental["Identifier"])

trees[mask]
```

[175]:

	Identifier	Number Of Trees	Site Name
66	00059712	1.0	Maiden Lane Estate
125	00048578	1.0	BUCK STREET
1148	00006577	1.0	FAWLEY ROAD
1998	00007366	1.0	FORTUNE GREEN RD, OPEN SPACE (LS)
2246	00014633	1.0	Mortimer Estate
5478	00060382	1.0	SHAFTESBURY AVENUE
10637	00002874	1.0	BURGHLEY ROAD
10977	00055227	1.0	BURGHLEY ROAD
11795	00016702	1.0	RED LION SQUARE, GARDENS (LS)
11856	00054744	1.0	Carrol & Sanderson Close Estate
12056	00003694	1.0	Estate 1-161 Burnham (cons)

12936	00054558	1.0	ST. MARY'S KILBURN C OF E JMI (E)
13248	00059317	1.0	ADELAIDE ROAD NATURE AREA
16815	00055884	1.0	HONEYBOURNE ROAD
18690	00059963	1.0	Ampthill Square Estate
18958	00059246	1.0	Belsize nature reserve, Russell Nursery
19606	00005127	1.0	CUMBERLAND MARKET, OPEN SPACE (LS)
20169	00017912	1.0	SHERRIFF ROAD
20226	00047080	1.0	Ampthill Square Estate
21287	00029059	1.0	Estate 1-20 Marrick House (cons)
22470	00012126	1.0	KINGS COLLEGE ROAD
23301	00010784	1.0	Holly Lodge Estate
23315	00056485	1.0	WATERLOW PARK (LS)

	Contract Area	Scientific Name	\
66	Housing	Acer saccharinum	
125	Highways	Sorbus aucuparia	
1148	Highways	Tilia euchlora	
1998	Parks	Ilex aquifolium	
2246	Housing	Tilia europaea	
5478	Highways	Vacant Tree Pit	
10637	Highways	Platanus x hispanica	
10977	Highways	Amelanchier lamarckii	
11795	Parks	Platanus x hispanica	
11856	Housing	Prunus unidentified species	
12056	Housing	Acer platanoides	
12936	Education	Amelanchier lamarckii	
13248	Parks	Stump Only	
16815	Highways	Acer pseudoplatanus 'Brilliant	
18690	Housing	Vacant Tree Pit (planned: Parrotia persica van…	
18958	Parks	Ulmus procera	
19606	Parks	Platanus x hispanica	
20169	Highways	Tilia platyphyllos	
20226	Housing	Malus unidentified species	
21287	Housing	Sambucus nigra	
22470	Highways	Fraxinus excelsior	
23301	Housing	Ilex aquifolium	
23315	Parks	Fraxinus excelsior	

	Inspection Date	Inspection Due Date	Height In Metres	Spread In Metres	\
66	2019-05-28	2022/2023	12.0	5.0	
125	2017-07-19	2020/2021	6.0	2.0	
1148	2018-09-28	2021/2022	15.0	6.0	
1998	2017-03-21	2019/2020	9.0	6.0	
2246	2019-01-29	2021/2022	16.0	12.0	
5478	NaT	NaN	NaN	NaN	
10637	2017-08-14	2020/2021	20.0	8.0	
10977	2017-08-14	2020/2021	3.0	2.0	

11795	2018-06-04	2021/2022	30.0	23.0
11856	2017-01-06	2020/2021	3.0	3.0
12056	2018-04-17	2021/2022	3.0	1.0
12936	2019-10-07	2022/2023	4.0	2.0
13248	2019-01-31	2021/2022	0.0	5.0
16815	2018-05-10	2021/2022	2.0	1.0
18690	2019-01-08	2022/2023	NaN	NaN
18958	2019-01-29	2021/2022	5.0	4.0
19606	2018-03-13	2020/2021	10.0	8.0
20169	2018-10-09	2021/2022	9.0	5.0
20226	2019-01-08	2022/2023	5.0	3.0
21287	2018-06-19	2021/2022	6.0	6.0
22470	2018-07-13	2021/2022	18.0	12.0
23301	2017-06-14	2020/2021	7.0	5.0
23315	2019-05-24	2022/2023	12.0	5.0

	Diameter In Centimetres At Breast Height	Ward Code
66	20.0	E05000131
125	10.0	E05000130
1148	38.0	E05000145
1998	44.0	E05000132
2246	47.0	E05000140
5478	NaN	E05000138
10637	52.0	E05000139
10977	5.0	E05000139
11795	165.0	E05000138
11856	13.0	E05000137
12056	6.0	E05000128
12936	8.0	E05000140
13248	50.0	E05000128
16815	4.0	E05000145
18690	NaN	E05000143
18958	11.0	E05000134
19606	38.0	E05000142
20169	42.0	E05000145
20226	16.0	E05000143
21287	39.0	E05000140
22470	48.0	E05000128
23301	20.0	E05000137
23315	16.0	E05000137

	Ward Name	Easting	Northing	Longitude
66	Cantelowes	529795	184142	-0.129750
125	Camden Town with Primrose Hill	528900	184020	-0.142698
1148	West Hampstead	525572	185015	-0.190313
1998	Fortune Green	525074	185541	-0.197304
2246	Kilburn	525763	183613	-0.188060

```
5478        Holborn and Covent Garden   530073    181247  -0.126813
10637                    Kentish Town   529119    185889  -0.138863
10977                    Kentish Town   528920    185694  -0.141802
11795       Holborn and Covent Garden   530572    181701  -0.119460
11856                        Highgate   528661    185556  -0.145585
12056                         Belsize   527015    184315  -0.169763
12936                         Kilburn   525443    183919  -0.192552
13248                         Belsize   527577    184297  -0.161664
16815                  West Hampstead   525593    185050  -0.189987
18690      St Pancras and Somers Town   529216    183103  -0.138479
18958                      Gospel Oak   527523    185233  -0.162101
19606                  Regent's Park   528913    182814  -0.142959
20169                  West Hampstead   525265    184543  -0.194897
20226      St Pancras and Somers Town   529279    183106  -0.137573
21287                         Kilburn   525832    183583  -0.187075
22470                         Belsize   526999    184416  -0.169957
23301                        Highgate   528472    186812  -0.147845
23315                        Highgate   528730    187264  -0.143955

        Latitude                Location
66     51.541266   (51.541266, -0.12975)
125    51.540371  (51.540371, -0.142698)
1148   51.550062  (51.550062, -0.190313)
1998   51.554901  (51.554901, -0.197304)
2246   51.537424   (51.537424, -0.18806)
5478   51.515186  (51.515186, -0.126813)
10637  51.557120   (51.55712, -0.138863)
10977  51.555408  (51.555408, -0.141802)
11795  51.519146   (51.519146, -0.11946)
11856  51.554226  (51.554226, -0.145585)
12056  51.543451  (51.543451, -0.169763)
12936  51.540245  (51.540245, -0.192552)
13248  51.543163  (51.543163, -0.161664)
16815  51.550377  (51.550377, -0.189987)
18690  51.532058  (51.532058, -0.138479)
18958  51.551587  (51.551587, -0.162101)
19606  51.529527  (51.529527, -0.142959)
20169  51.545886  (51.545886, -0.194897)
20226  51.532073  (51.532073, -0.137573)
21287  51.537139  (51.537139, -0.187075)
22470  51.544363  (51.544363, -0.169957)
23301  51.565562  (51.565562, -0.147845)
23315  51.569567  (51.569567, -0.143955)
```

[176]:
```
# Confirmed how many rows we have
```

```
trees[mask].shape
```

[176]:
```
(23, 17)
```

0.7.2 8.2 Find Environmental Data that Doesn't have Matching Tree Data

[184]:
```
# Found environmental data that doesn't have matching tree data

mask = ~environmental["Identifier"].isin(trees["Identifier"])
environmental[mask]
```

[184]:
```
Empty DataFrame
Columns: [Identifier, Maturity, Physiological Condition, Tree Set To Be Removed,
Removal Reason, Capital Asset Value For Amenity Trees, Carbon Storage In
Kilograms, Gross Carbon Sequestration Per Year In Kilograms, Pollution Removal
Per Year In Grams]
Index: []
```

[185]:
```
# Confirmed how many rows we have

environmental[mask].shape
```

[185]:
```
(0, 9)
```

0.7.3 8.3 Find Trees that Don't have Matching Common Names Data

[179]:
```
# Found trees with scientific names that don't have matching common names data

mask = ~trees["Scientific Name"].isin(names["Scientific Name"])

trees[mask]
```

[179]:
```
       Identifier  Number Of Trees                              Site Name  \
151      00051832              1.0                             ARGYLE WALK
384      00053954              1.0                          CHURCHILL ROAD
495      00047497              1.0                          PATSHULL PLACE
611      00055434              1.0                   SHARPLES HALL STREET
653      00055289              1.0                      QUEEN'S CRESCENT
…           …                  …                                      …
21826    00050835              1.0                             INGESTRE RD
22948    00052341              1.0                    NEW COMPTON STREET
23266    00048846              1.0                             ASMARA ROAD
23335    00048705              1.0                       GOLDINGTON STREET
23372    00031627              1.0  ST. GEORGE THE MARTYR C OF E JMI (E)

      Contract Area                Scientific Name Inspection Date  \
151        Highways  Sorbus aucuparia 'Streetwise'      2019-02-10
```

```
384        Highways  Sorbus aucuparia 'Streetwise'      2017-10-07
495        Highways  Sorbus aucuparia 'Streetwise'      2017-06-22
611        Highways  Sorbus aucuparia 'Streetwise'      2019-09-30
653        Highways  Sorbus aucuparia 'Streetwise'      2017-07-08
…             …                              …             …
21826      Highways  Sorbus aucuparia 'Streetwise'      2017-08-18
22948      Highways  Sorbus aucuparia 'Streetwise'      2019-07-08
23266      Highways  Sorbus aucuparia 'Streetwise'      2018-08-28
23335      Highways  Sorbus aucuparia 'Streetwise'      2019-10-23
23372     Education       Cotoneaster salicifolius      2018-07-23

      Inspection Due Date  Height In Metres  Spread In Metres  \
151             2022/2023               7.0               3.0
384             2020/2021               3.0               2.0
495             2020/2021               5.0               3.0
611             2022/2023               2.0               2.0
653             2020/2021               4.0               1.0
…                   …                 …                 …
21826           2020/2021               3.0               1.0
22948           2022/2023               4.0               3.0
23266           2021/2022               5.0               3.0
23335           2022/2023               6.0               2.0
23372           2021/2022               5.0               5.0

       Diameter In Centimetres At Breast Height  Ward Code  \
151                                        12.0  E05000141
384                                         5.0  E05000139
495                                        11.0  E05000131
611                                         4.0  E05000130
653                                         7.0  E05000136
…                                           …          …
21826                                       7.0  E05000139
22948                                       6.0  E05000138
23266                                       8.0  E05000132
23335                                      12.0  E05000143
23372                                       8.0  E05000138

                          Ward Name  Easting  Northing  Longitude  \
151                     King's Cross   530227    182706  -0.124054
384                     Kentish Town   529007    185975  -0.140440
495                       Cantelowes   529202    184717  -0.138094
611    Camden Town with Primrose Hill   527962    184050  -0.156202
653                       Haverstock   528072    184723  -0.154369
…                                ...      …         …          …
21826                   Kentish Town   528962    185826  -0.141137
22948      Holborn and Covent Garden   529976    181160  -0.128249
23266                  Fortune Green   524568    185347  -0.204661
```

```
23335      St Pancras and Somers Town   529662    183417  -0.131941
23372       Holborn and Covent Garden   530742    182119  -0.116856

        Latitude                Location
151    51.528257  (51.528257, -0.124054)
384    51.557913   (51.557913, -0.14044)
495    51.546569  (51.546569, -0.138094)
611    51.540853  (51.540853, -0.156202)
653    51.546881  (51.546881, -0.154369)
…           …                       …
21826  51.556589  (51.556589, -0.141137)
22948  51.514422  (51.514422, -0.128249)
23266  51.553268  (51.553268, -0.204661)
23335  51.534780   (51.53478, -0.131941)
23372  51.522865  (51.522865, -0.116856)

[76 rows x 17 columns]
```

[181]:
```
# Confirmed how many rows we have
trees[mask].shape
```

[181]:
```
(76, 17)
```

0.7.4 8.4 Observations

For the Identifiers in the Trees data and Environmental Dataset, there were 23 rows that didn't contain matching data.

For the Environmental data and the Trees data, I didnt identify any mismatches between the data in terms of the Identifier.

For the Trees data and Common names data, I identified 76 rows that didn't match in terms of the Scientific name.

1 END OF NOTEBOOK

[ ]:
```

```