# Data Requirements

**Requirement**: Brief name of the requirement.

**Data File**: Name of data files or other source

**Additional Data Needed**: Any additional data that is needed to meet the requirement. I.e. data that is not available in the supplied data files.

**Usage:** How will the data be used in the project to meet the requirement? E.g. will data be combined? Will it be fed into another tool? Will it be sent somewhere else?

|  |  |  |  |
| --- | --- | --- | --- |
| **Requirement** | **Data File** | **Additional Data Required** | **Usage** |
| Tree Walks Brochure | trees | Missing Values : Have a relatively low percentage of missing values (below 3%).  Outliers : remove small percentage (>0.009%) of outliers to minimize the effect in three columns (Height, Spread, Breast Height)  Duplicates: 6 duplicates are found in the column ‘Identifier‘  Geolocation issues: 56 rows are removed due to null or 0 geo-coordinates with the Easting and Northing columns. And if only selecting the rows containing trees inside of Camden, 17073 rows are left and 6315 rows are removed.  Unmatch : Need 23 trees (0.1%) from the environmental dataset and 76 trees (0.3%) from the names dataset | Provide main infromation about trees.  Report in PowerBi/Tablau  Data Analysis Project |
| Environment Report | environmental | Missing Values : Have a relatively low percentage of missing values (below 3%). Except three columns (Carbon Storage, Gross Carbon Sequestration, and Pollution Removal) have around 12.21% missing values  Missing Values : The column "Removal Reason" in the Environmental dataset has a very high percentage of missing values (99.64%), which suggests that most of trees do not need be removed.  Unmatch : 23 trees with column ‘Identifier’ | Provide environmental data about each tree, combine with the trees dataset with column ‘Identifier’ |
| Public Tree Data | names | Missing Values : Column ‘Common Name’ has a relatively low percentage of missing values (around 4%).  Unmatch : 76 trees with column ‘scientific names’ | Provide tree (common) names data, combine with the trees dataset with column ‘scientific names’ |