Data Requirements

Use this template to record how the requirements are met by the data.

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	- camden_trees.xlsx -tree_common_names.json	**Clean Trees Dataset:** **Completeness** : for Qualitative Continuous column as(Height in Metres, Spread In Metres, Diameter In Centimetres At Breast Height) we can impute missing by mean - Enter missing data in location 0.2%. If we couldn't manage to find it we can delete it. - Enter missing data Inspection Date 1.7%, or Impute data with Mod. - Qualitative Nominal, we can fill missing value by 0.9% with "null" or impute with mod **Accuracy:** Delete Duplication 0.05% - Handel Outlier in location 0.05%	- Data Combined - Data clean by python - Data organise in database - Feed tableau for visualisation & report

		Based we have the correct boundary of Camden Council,If we know correct location from council otherwise We can delete it -Handle outlier in tree specification, We need the original specification max and min for trees to make sure of the outlier and we have the option to impute data with the mean. **Clean Common Names Data:** **Completeness** : missing 4% of common names. Search for another source to scrap missing data and complete it. Try to build data for Common Name internal ownership. **Provide more information:** - About the characteristics standar of the trees such as what is the max and min height, Spread and Diameter for each category. -Actual box boundary for Camden council (for measuring outlier with more accurately) -Another source for Common Name	
Environment Report	-camden_trees_enviro nmental.csv -tree_common_names .json	**Matching Tree Data for Environmental Records:** Similarly, there are entries in the environmental dataset that do not have matching tree data. To ensure consistency and accuracy, acquiring or linking tree data for these environmental records is essential. We have a big challenge here: 99% of data didn't match. We need to make sure consistency in format and set plans to fix this mismatch.	- Data Combined - Data clean by python - Data organise in database - then feed data into visualisation tool as Tableau - Generate requirement analysis - Display visualisation on to website

		Complete Environmental Data: The environmental dataset seems to be missing some crucial environmental attributes for certain trees, such as carbon storage, carbon sequestration, and pollution removal. Acquiring complete environmental data for all trees could provide a comprehensive understanding of their environmental impact. -Fill missing data for Quantitative continuous 12% by impute data with mean not recommended because percentage high. Prefer to enter this data as possible to reduce this missing percentage then impute the rest.	
Public Tree Data	-tree_common_names .json -camden_trees_enviro nmental.csv - camden_trees.xlsx -tree_common_names .json	**Integration:** integration with other datasets (e.g.environmental data, common names data), ensure that the trees dataset can be seamlessly integrated with these sources to enhance the quality and richness of the information provided. **Matching Tree Data for Environmental Records:** Similarly, there are entries in the environmental dataset that do not have matching tree data. To ensure consistency and accuracy, acquiring or linking tree data for these environmental records is essential **Recommended:** conducting data validation checks, verifying the accuracy of key information, and ensuring that the dataset meets the specific requirements and objectives of the initiatives.	- Data Combined - Data Clean - Data organise into database - upload link to the data in to website with data