

# TREE ANALYSIS: CU BOULDER (BUFFALIZATION DATA VISUALIZATION CHALLENGE)

---

## 1. Data Cleaning Process

The dataset provided for this challenge contained information on **777 trees** located in **CU-Boulder's historic campus**. To ensure accuracy and reliability in analysis, the following data cleaning steps were performed:

### 1.1 Handling Missing Values

- Conducted a thorough assessment of missing values in key fields such as **Canopy Spread, Height, and Species**.
- Applied **mean imputation** for numerical fields where missing values were minimal to maintain dataset integrity.
- Removed records containing critical missing data that could not be reasonably inferred, ensuring high-quality insights.

### 1.2 Removing Duplicates

- Performed a **duplicate check** based on **Species, Height, Canopy Spread, and Location Coordinates** to identify redundant records.
- Eliminated duplicate entries to maintain data accuracy and prevent skewed analysis.

### 1.3 Outlier Detection & Removal

- Applied the **z-score method** to detect anomalies in **Height and Canopy Spread** values.
- Conducted manual verification to distinguish between actual extreme values and data entry errors.
- Removed incorrect outliers while preserving biologically valid data points for accurate analysis.

A refined dataset, free from inconsistencies, was finalized and saved as **cleaned\_trees\_dataset.csv** for further analysis.

---

## 2. Calculated Fields

To enrich the dataset and derive meaningful insights, several calculated fields were introduced. These fields provide deeper understanding into tree structure, growth dynamics, and environmental impact.

## 2.1 Canopy Area

- **Formula:**  $\pi \times (\text{Canopy Spread} / 2)^2$
- Measures the total ground area covered by a tree's canopy, aiding in understanding its environmental footprint.

## 2.2 Canopy Volume

- **Formula:**  $(2/3) \times \pi \times (\text{Canopy Spread} / 2)^3$
- Provides an estimate of the total airspace occupied by a tree's canopy, assuming a semi-spherical shape.

## 2.3 Canopy Density

- **Formula:**  $\text{Canopy Volume} / \text{Canopy Area}$
- Determines the compactness of the tree's canopy, offering insight into foliage density.

## 2.4 Canopy Height Ratio

- **Formula:**  $\text{Canopy Spread} / \text{Height}$
- Assesses the proportional relationship between a tree's canopy width and its overall height.

## 2.5 Height Canopy Ratio

- **Formula:**  $\text{Height} / \text{Canopy Spread}$
- Evaluates how tall a tree stands in relation to the spread of its canopy.

## 2.6 Bio Mass

- **Formula:**  $\text{Height} \times \text{Canopy Spread}$
- Provides an approximation of a tree's total biomass, contributing to ecological impact assessments.

## 2.7 Carbon Absorbed

- **Formula:**  $\text{Bio Mass} \times 0.5$
- Estimates the amount of carbon a tree absorbs, assuming 50% of biomass consists of carbon.

## 2.8 Slenderness Index

- **Formula:**  $\text{Height} / \sqrt{\text{Canopy Spread}}$
- Identifies trees with potential structural instability by measuring their height-to-canopy ratio.

## 2.9 Growth Stage Classification

### Formula:

```
IF [Height] <= 20 THEN "Sapling"  
ELSEIF [Height] <= 35 THEN "Maturing"  
ELSEIF [Height] <= 47 THEN "Mature"  
ELSE "Ancient"  
END
```

- Trees are categorized into **Sapling, Maturing, Mature, and Ancient**, offering insights into growth distribution across campus.
- 

## 3. Storytelling & Insights

The trees on CU Boulder's campus play a crucial role beyond aesthetics, contributing significantly to **carbon sequestration, environmental sustainability, and ecological diversity**. This dashboard presents a **data-driven narrative** that highlights:

### Canopy Coverage & Carbon Sequestration

- Trees with **expansive canopies** contribute significantly to **carbon absorption**, reducing environmental impact.
- The **Canopy Coverage Map** illustrates tree density across campus, highlighting areas of concentrated foliage.

### Growth Stages & Structural Insights

- The **Growth Classification Chart** reveals that a substantial proportion of trees fall within the *Maturing and Mature* categories, supporting long-term environmental benefits.
- The **Slenderness Index Analysis** identifies trees with potential structural vulnerabilities, assisting in proactive maintenance efforts.

### Tree Diversity & Distribution

- The **Tree Distribution Map** visually represents the spread of different species across campus, emphasizing biodiversity.
- The **Top N Analysis** showcases trees with **exceptional canopy height ratios and superior carbon absorption rates**, underscoring their ecological significance.

By integrating **scientific analysis with compelling visuals**, this dashboard effectively conveys the environmental importance of CU Boulder's trees while fostering a deeper appreciation for urban forestry.

---

## 4. Conclusion

This analysis provides a **holistic understanding of CU Boulder's trees**, highlighting their ecological value, structural attributes, and carbon sequestration potential. The cleaned dataset and calculated fields facilitated an in-depth exploration of tree-related metrics, while the dashboard offers an engaging visual narrative of these findings.

This project stands at the **intersection of data science, sustainability, and storytelling**, shedding light on the often-overlooked impact of campus trees in a manner that is both informative and visually compelling.

---

**Author: Trinay Gangisetty**

*Buffalization Data Visualization Challenge 2025*