DATA 601 – Project Proposal

Exploring the City of Calgary's Public Trees Dataset

September 30th, 2019

Michael Ellsworth

Jennifer Panek

Christy Sarmiento

Introduction

Trees within municipalities have proven benefits. By improving the aesthetics of communities and filtering the air we breathe, trees bring a variety of quantifiable advantages to cities that are not typically studied.

In Calgary, there are approximately 7 million trees within city limits. What makes this statistic impressive is that Calgary's arid, prairie type climate is not necessarily conducive to growing trees, let alone 7 million of them. This becomes a difficult challenge for the local government because in large part, trees are a publicly owned asset and the cost and resources associated with the maintenance of those trees fall on the municipality. Additionally, understanding the value of investing resources into trees is not particularly intuitive for citizens.

As pressure on the City to demonstrate fiscal restraint continues to increase, the budget for maintaining public trees is unlikely to grow. The purpose of this project is to analyze the City of Calgary's public tree data to come up with potential recommendations that may assist in improved cost and resource allocation that allows the City to maintain or improve the condition of their public trees. This analysis will also attempt to present helpful insight into the value that trees provide municipalities so that citizens can better understand the importance of investing their tax dollars into this shared asset.

Dataset

The City of Calgary has a publicly available dataset that provides a variety of information on Calgary's trees including species type, maintenance level, breast height diameter, overall rating, condition rating, heritage tree designation and location. This is a structured dataset with recorded information on almost 500,000 trees. This information was collected by trained Urban Forestry staff who visit each tree and populate the information accordingly.

The Public Trees[1] dataset is from the Open Calgary data portal. This dataset was created on January 18th, 2018 and is updated on a weekly basis. In total, this dataset includes 496,000 observations and 20 variables. Each observation represents a single tree and each variable represents information on that specific tree. The variables in this dataset are primarily categorical with a few numerical variables.

The data is provided by the City of Calgary and contains information licensed under the Open Government License – City of Calgary[2]. A link to the license agreement is provided in the References section and grants us permission to use this data for our analysis.

Guiding questions

As previously mentioned, the City of Calgary has approximately 7 million trees within city limits. As this dataset has information for under 500,000 of them, our group intends to begin by exploring the quality of the dataset and answer the following questions:

- Are there noticeable gaps in the data?
- Is there anything about the missing data that would help provide context for further analysis?

Following the data quality check, our group intends to explore how tree density varies across the city. This step helps set the stage for further analysis by getting a sense of where trees are located and which

regions of the city host certain types of trees. This analysis can be linked to the other analyses described below. The questions that will guide this analysis include:

- What does the general tree density look like across the city?
- Are there noticeable differences in the density of different types of trees?
- Is it apparent where the highest density of good condition trees resides?

Next, the group is interested in determining how different variables affect the condition and maintenance level of each tree. This analysis is relevant to the discussion on how the City allocates its resources on public trees and may help determine areas of optimized resource allocation. The questions that will be asked here include:

- Do certain types of species require more maintenance?
- Does size affect the condition of trees?

Lastly, we will attempt to quantify the value of the trees in Calgary in terms of their ability to capture carbon from the atmosphere and prevent storm water runoff. This analysis intends to help justify the cost and effort that the City invests into public trees. Based on these two categories, we will attempt to answer the following questions:

- How much value does each tree provide?
- What is the density of the value in each community?

Tasks

The quality check step of this project will focus on grouping data by each variable and assessing the counts of the missing data in each column. This step will also be a good opportunity to check the formatting of the data to ensure it is consistent and clean. The missing data will be visualized simply through bar charts in order to quickly asses the quantity of missing data for each variable. Christy Sarmiento will be responsible for the tasks outlined in this section.

For the tree density analysis, “groupby” and “agg” functions will be used to determine tree counts per variable. Additional region specifiers from other public datasets will be joined onto this dataset as well. Tree density may be visualized on bar charts or a map using GeoPandas, choropleth maps in matplotlib and plotly. Jennifer Panek will be responsible for the tasks outlined in this section.

For the condition and maintenance level analysis, “groupby” and “agg” functions will again be used to evaluate condition and maintenance per variable. Histograms will be used to visualize the counts of each variable. Michael Ellsworth will be responsible for the tasks outlined in this section.

Research will be done to help quantify the value of certain trees in terms of carbon capture capacity and storm water avoidance. Based on preliminary research, it is not clear if this data is publicly available but if possible, two new columns will be created based on these researched values. As the data is not available in the current dataset and this group is unsure if the data is available to complete this task, the task will be completed as a group following the completion of the other tasks.

The aforementioned tasks will use the “pandas” library for all indexing, filtering, sorting and function application. Visualizations will be completed using the “matplotlib” library.

References

1 City of Calgary (2019) *Public Trees* [Online]. Available at: https://data.calgary.ca/Environment/Public-Trees/tfs4-3wwa (Accessed: 28 September 2019)

2 City of Calgary (2019) *Open Government Licence - City of Calgary* [Online]. Available at: https://data.calgary.ca/stories/s/Open-Calgary-Terms-of-Use/u45n-7awa (Accessed: 28 September 2019)