NEW YORK STREET TREE CENSUS 2015 DATA ANALYSIS

PRESENTED BY: **GROUP 6**

MASTER OF PROFESSIONAL STUDIES IN ANALYTICS

ALY6070 21601: Communications and Visualization for Data Analytics

March 9, 2022

To: **PROF. MYKHAYLO TRUBSKYY**

Understanding the data

The New York city's Parks and Recreation department conducted a census of over 600k trees across New York city. The campaign leveraged high-tech tools and mapping techniques to capture information related to the tree species, measurements, health, problems, and location. To answer the questions and generate actionable insights, we will use the following **variables**: -

- **Tree measurements** – Diameter of trees and stumps ('tree_tbh', 'stump_diam', etc.)
- **Health and Problems of the trees** – Qualitative variables indicating the health, status, and the problems caused to the trees due the surroundings ('health', 'status', 'curb_loc', 'sidewalk', 'steward', 'root_stone', 'trunk_wire', 'brch_light', etc.)
- **Location of the tree** – Location attributes of the tree to understand plant coverage ('zipcode', 'boroname', 'latitude', 'longitude', etc.)

Initial analysis of the variables

To design the dashboard, and communicate insights from the data, a thorough understanding of data is required. In the following plots, we tried to understand the distribution of some of the important variables like '*health*', '*status*', and '*tree diameter*'.
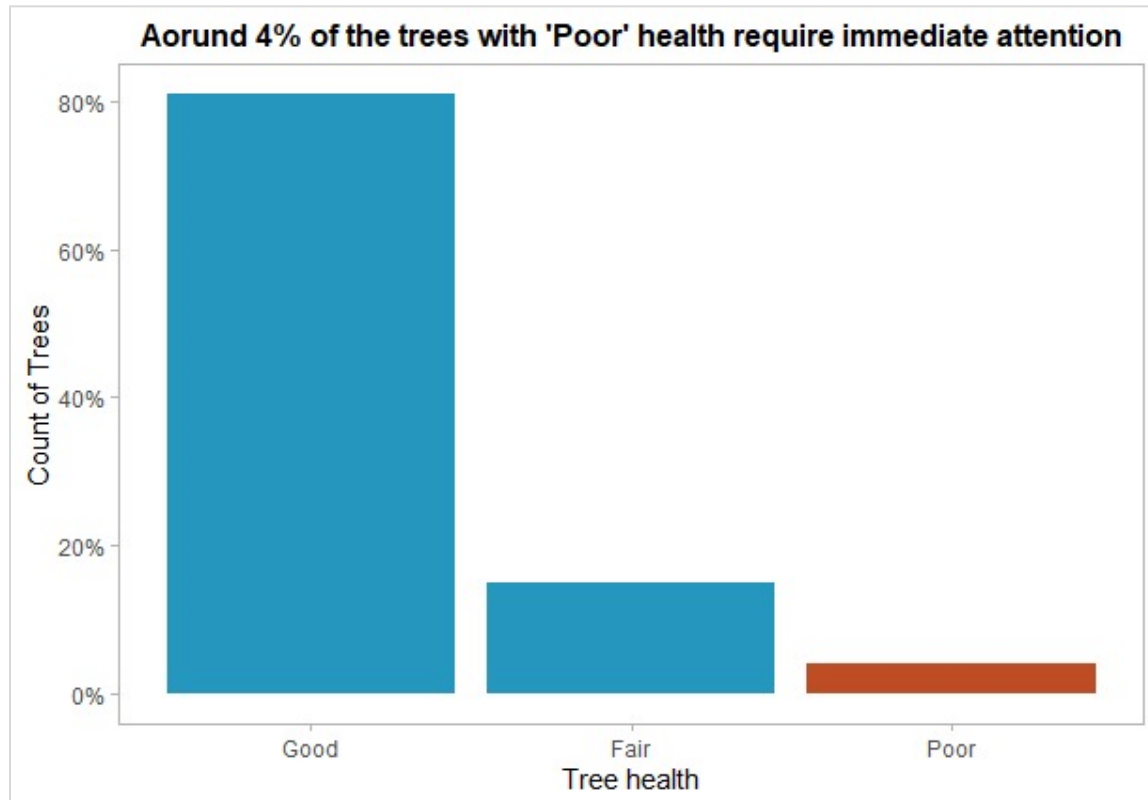
Observation
- A small fraction of trees has 'Poor' health which require assistance from the volunteers and the New York City department
  - We can use this variable to identify the major species and neighbourhoods that require immediate attention

Observation
- Trees with better health have higher diameter
  - This variable needs to be further analysed and visualized to segregate tree species that have lesser diameter and trees with 'poor' health that have lesser diameter
- *Note -* 'Dead' trees and 'tree stumps' have been eliminated from this analysis

Audience of the Data

1. **Municipal Urban Forest Management:** The group can effectively utilize the data for planning and providing care and management trees in urban setting.
2. **Citizen Engagement Team:** This data can be leveraged by government engagement teams to form interactions with citizens and plan to work collaboratively to implement policies.
3. **Department of City Planning:** Tree census data can be used by this department to improve urban life. It can be used to carry out numerous activities like establishment of public parks, planning for long term city management process etc.

Questions to the Dataset

1. What are the overall statistics of the tree plantation and urban forestry in New York?
2. Identify the top tree species in New York and any problems related to them.
3. Can we track tree survival and health status across the city?
4. What are the major causes that incite problem in tree parts?
5. What are the top species that require guard replacements?
6. Can we identify top boroughs with the highest count of stumps for replantation and tracking illegal cutting of trees?
7. What are the top boroughs with the highest count of dead trees (to take caution for wildlife habitat)?
8. What is tree cover density across all neighbourhoods of New York?

Visually answerable questions

A Data Dashboard is an information management tool used to track, analyse, and display key performance indicators, metrics, and data points. A dashboard can be used to monitor the overall health of businesses, departments, or a specific process.

We will use dashboard to show the data clearly and perform various operations on it:

1. Dashboard will be used to aggregate data from multiple data sources in an automated manner (if needed be).
2. It wouldn't just save time, but enable us to visualise and analyse all of the data together.

3. Dashboard reporting can have effect on internal and external relationships. It is consistent and cross-functional communication.
4. Trend Analysis can be conducted using dashboards visually to check what's worked (or what hasn't) in the past to shift approach for the future.
5. It would help to monitor multiple metrics and KPIs (Key Performance Indicators) at once.
6. It could enable to customize the visualizations to highlight the story-telling points and summarise the data pool without any textual information.

Types of graphs and charts to use and answer the business questions

"A Picture Is Worth a Thousand Words" - Fred R. Barnard (1921)

Certainly, when it comes to data, metrics, and KPIs, nothing could be more apt. Without visual tools such as charts and graphs we'd soon get lost in the wind of numbers being presented to us. These tools simply take that data and put it into a concise and understandable format.

There are breadths of charts, graphs, reports available to us for dashboard reporting but according to our initial impressions of the data, we will be going forward with the following charts and graphs -

1. *Bar Chart* - It will enable us to analyse the trees count in different boroughs and answer questions related to it.
2. Side-by-side view of *Bar Chart* (coordinate-flipped) and *Heat Map*
   Analysis of the trees in the council districts or neighbourhoods within a borough.
3. *Stacked Bar Chart* - Tree Health will be analysed using this chart type for better understanding.
4. *Line Chart* - To visualise the diameter distribution in the plants' species and investigate any patterns.

References

1. *TreesCount! 2015 : NYC Parks*. (2015). NYC Parks.
   https://www.nycgovparks.org/trees/treescount

2. What is a data dashboard? (2021). Klipfolio. https://www.klipfolio.com/blog/what-is-a-data-dashboard