BU Sustainability: Water Bottle Stations

This repository contains a software pipeline to determine new locations to install water bottle stations. There are three tasks this project aims to accomplish:

1. Establish how many people are impacted by each water filling station in the current inventory
2. Establish a list of factors and criteria for installing new stations
3. Provide a list of buildings to install new stations by applying the factors and the number of people that would have access to each of them

Overview

The BU Sustainability team needed assistance in identifying locations for new water bottle stations on the Charles Campus of Boston University based on various factors, explained under “Our Accomplishments”. Building stations at every facility is neither cost-effective nor practical, so our mission was to select sites that would yield the highest benefits. Our analysis led to the selection of 16 optimal locations for the stations, ranked in order of priority. Since the client did not specify the number of stations needed, we provided a list of 16 prioritized stations, enabling the client to implement the water bottle stations in the recommended sequence.

The top three prioritized buildings are:

736 - 738 Commonwealth Avenue (benefits 692 people)
514-522 Park Drive (benefits 598 people)
232 Bay State Road (benefits 720 people)

Following the top three buildings, the next 13 stations are listed in order of priority. If the clients decide to construct 10 water bottle stations, they should start with the three previously mentioned locations and proceed with the first seven buildings from the list below.

10 Lenox Street (benefits 659 people)
949 Commonwealth Avenue (benefits 830 people)
213 Bay State Road (benefits 1511 people)
580 Commonwealth Avenue (benefits 845 people)
619 Memorial Drive (benefits 1067 people)
910 Commonwealth Avenue (benefits 1607 people)
540 Commonwealth Avenue (benefits 563 people)
10 Buick Street (benefits 1081 people)

15 St. Mary's Street (benefits 1868 people)
975 Commonwealth Avenue (benefits 1054 people)
677 Beacon Street (benefits 667 people)
602 Commonwealth Avenue (benefits 829 people)
163-179 Amory Street (benefits 1359 people)

Notice that we did not simply rank the buildings based solely on the number of potential users for each station. We considered additional factors to establish the priority list, as detailed in the "Our Accomplishments" section below and further explained in the report.

Our Accomplishments

Our team received six datasets from the clients, including the BU Map of Water Filling Stations, Inventory of known/existing filling stations, Inventory of campus buildings and rooms with kitchen, BU Wifi Data for foot traffic, BU floor descriptions, and BU high volume event spaces. We carefully examined each dataset and refined them to focus solely on key factors. For example, crucial data features extracted from the Wi-Fi data were date, hour, building_desc, latitude, longitude, and capacity.

Initially, we identified six potential factors to determine the optimal locations for new water bottle filling stations: kitchen availability, air ventilation availability, maintenance, high density, foot traffic, and cost. However, the high-density dataset only contained four locations outside of the Charles Campus, allowing us to disregard it. Furthermore, the client specified that installation costs should not influence our decisions, but requested a list of recommendations, leaving us with four primary factors to guide the selection.

Since water filling stations are more beneficial in buildings with higher occupancy, foot traffic data was the most crucial data to begin our search. We established a threshold value for existing water bottle stations based on the median daily user count and applied this benchmark to buildings without stations, narrowing our list down to 20 buildings.

Among these 20 buildings, we considered kitchen availability, maintenance, and air ventilation availability into account. However, every building in the Wifi dataset already had kitchens, forcing us to exclude this criterion from our calculations. Four of the 20 buildings lacked BU maintenance, a crucial factor given the need for regular maintenance of station filters. After deleting these four buildings, three of the 16 buildings lacked mechanical ventilation. Therefore, these three buildings emerged as the top priorities for new water bottle stations.

After discussion with the clients, they recommended ranking all remaining 13 buildings by their importance. We applied an additional criterion: distance to existing water bottle stations in a special format. Buildings closer to existing water bottle stations were considered less significant as they are conveniently accessible for refilling bottles. We scored each of the 13 buildings based on our algorithm, with higher scores indicating higher priority.

This approach allowed us to establish a specific order of importance for the remaining 13 buildings, detailed in the "Summary" section above.

Navigating Files in the Repo

In the GitHub Repository on the "team-e" branch, we organized the files in a clear and structured manner. Inside the folder "sp24-team-e" folder, a few files offer an overview of the project and our team.

./data

The "data" folder within "sp24-team-e" contains six datasets we utilized to identify possible locations to install water bottle stations and includes Python analysis of key features. The raw datasets are also accessible within the "BU Sustainability: Water Bottle Stations" folder on Google Drive. As explained above, we used six datasets provided by the clients. They are

BU Map of Water Filling Stations
Inventory of Known/Existing Filling Stations
Inventory of Campus Buildings and Rooms with Kitchen,
BU Wifi Data for Foot Traffic
BU Floor Descriptions
BU High Volume Event Spaces

Each subfolder in the "data" folder contains the dataset and the analysis of the dataset to extract significant features. For every dataset, we examined the columns in detail and removed those that were irrelevant or had excessive missing values (NaNs). For example, most values in the "Accessibility" column in the Inventory of Known/Existing Filling Stations were NaN, which justified its elimination. We transformed some categorical data for significant columns into numerical format by assigning an integer to each unique value. Moreover, we generated histograms to thoroughly analyze significant columns in the Early Insight Report. For instance, we calculated the average daily occupancy per building and created histograms to display which buildings had the highest occupancy for the Wifi data.

./deliverables

The "deliverables" folder in "sp24-team-e" contains the three reports (the Early Insight Report, Mid-Semester Report, and the Final Report). In addition, it also holds Python scripts used to calculate the threshold values, application of the additional factors (kitchen, maintenance, air ventilation, and distance) to determine the list of buildings requiring new water bottle stations, and the map to show visualization of our final results.

How to Run Code/Reproduce the Results

Our project includes five Python notebook files. The "EarlyInsightReportNotebookCode.ipynb" in the deliverable/Early_Insight folder handles the cleaning and analysis of the datasets. The "CombinedWifiDataFilter.ipynb" file, located in the deliverable/mid_semester folder, assesses the average daily occupancy of each unique building from 7 AM to 2 AM based on the wifi usage. We used the daily wifi data to establish a threshold value based on the median, which helps narrow down the list of buildings without water bottle filling stations. Additionally, we applied the three factors (kitchen, mechanical ventilation, and maintenance) in the "Threshold_and_other_factors.ipynb" in the same folder. The "distance.ipynb" file was used to apply a distance factor for prioritizing the remaining 13 buildings and visualizations were created in the "map.ipynb" file.

Information Future Users Require

For future reference, it is important to note that we manually found the coordinates of the buildings to compute the Manhattan distance between them. To obtain the longitude and latitude pairs for each location, we used https://www.gps-coordinates.net/. These coordinates were then utilized within a Python dictionary to calculate the distances. Although a Python library could have been used for this purpose, we were unable to find the latitude and longitude pairs for specific street addresses, which led us to opt for this manual approach.

Furthermore, the clients reported that some existing water bottle filling stations were not listed in the dataset, indicating that the data was incomplete. According to them, only 80% of the stations exist in the dataset, implying that the buildings within our solution may already have a water bottle station. Therefore, we recommend running the codes once again after the dataset is fully updated.

References

1. Latitude and Longitude coordinates of the streets: https://www.gps-coordinates.net/