Early Insights Presentation

BU Sustainability: Understanding how weather impacts waste

Team 1: Junyi Zhu, Abdelazim Lokma, Zhengyu Liu, Karan Vombatkere
March 24, 2023

Project Overview

- BU Sustainability supports the transformation of Boston University's planning, operations, and culture toward a sustainable and equitable future.
- We have detailed data sets from each **monitor, overall data, temperature data, and waste generation spreadsheet by data from Casella**. The analysis of these datasets will inform BU Sustainability how they can potentially improve where to store waste if there are adverse weather effects.

Main project goals:

1. Combine the various different datasets from Casella and Contelligent, to ensure all the different data sources are being incorporated correctly.
2. Perform an analysis to answer the following questions that would help BU Sustainability implement a Zero Waste plan:
    a. Does temperature impact waste generation?
    b. If so, in what ways?
    c. Can temperature be used as a predictor of waste generation?

Team Progress

1. **Data Audit, Exploration and Combining sources**
    a. We explored different aspects of the dataset - PSI, temperature, waste, alerts, events, notifications, compactions, etc
    b. Performed a full data audit and preliminary analysis of each dataset
    c. Combined different datasets into comprehensive dataset for further analysis - join using device ID and serial no.

2. **Preliminary Analysis highlights**
    a. For each device, explore different patterns between waste generated, temperature and psi.
    b. Heatmap for each location. Shows Comm Ave produced the most trash (Charles River). There's more traffic on the BU Main Campus compared to other campuses
    c. Tonnage changes at popular sites - focus on the locations that have the most waste, and also explore the seasonality of waste generation at these sites
    d. Correlation between temperature and tonnage of Trash containers. Collected extra data from NOAA database and performed analysis
    e. Compactions by site - Warren towers and GSU should be considered separately
    f. Analysis of daily waste types - alerts

Data Audit and Combining sources

- Compiling multiple data sources:
  - Individual device readings - data, weather, PSI
  - Device Daily weights - Daily tonnage of waste by site and waste material type (trash, recycling, compost)
  - Alert flag history - alerts triggered by date and site
  - Hauler response, notifications and events - additional event data and notifications
  - Device compactions - total number of compactions by site, hourly breakdown

- Successfully audited and compiled temperature, PSI, and site for 21 Sites. Future work is to merge daily weight trash type and tonnage data for available date ranges.

- Challenges faced while joining data sources
  - Serial numbers do not match across files (2 different sets of serial numbers, as well as a device ID)
  - Site names aren't listed in a consistent manner
  - Date overlap between files is minimal - not enough overlapping dates to get a comprehensive overview with all features.
    - Min and max dates are offset by several months between files

```
BU #35 22 Babbit                        396
BU #82 Warren Towers 35                 394
BU #2 Student Village                   392
BU #90 - School of Law                  388
BU #87 College of Engineering           385
BU #46 30 Bay State                     382
BU MED - 15 Stoughton                   379
BU #43 West Loading Dock                377
BU #38 Life Sciences                    377
BU #72 Rafik B Hariri                   372
BU #48 Student Village #2               361
BU 140 Bay State Rd                     360
BU #4 Yawkey                            359
BU 685 Comm Ave                         356
BU #93 George Sherman Union             336
BU #96 808 Commonwealth                 335
BU Med 700                              279
BU #69 - Graduate Apartments            270
BU #105 Kilachand Hall                  240
BU #102 Student Health Services         236
BU Med #815 Albany                      192
Name: Site, dtype: int64
```

Data Merging Process

1. First concatenate all the readings files, and use the device filename as device_id
2. Extract a device_id to serial number mapping from Events file
3. Use extracted mapping to match PSI + weather readings
    a. Multilevel join using date + serial number
4. Join in Site information using serial number as a join key
5. Use Site name to join with Daily weights to incorporate tonnage
    a. Multilevel join using date + Site name

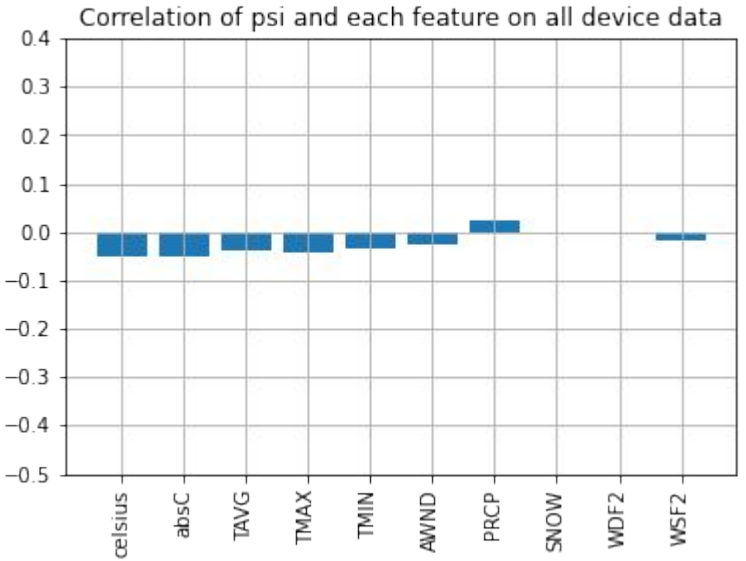
	device_id	serial_no	date	valuePsi	celsius	fahrenheit	Site
0	39671	31180	2022-07-01	253.333333	24.822500	76.680500	BU #82 Warren Towers 35
1	39671	31180	2022-07-02	331.200000	28.178000	82.720400	BU #82 Warren Towers 35
2	39671	31180	2022-07-03	280.400000	26.006000	78.810800	BU #82 Warren Towers 35
3	39671	31180	2022-07-04	279.272727	24.530909	76.155636	BU #82 Warren Towers 35
4	39671	31180	2022-07-05	182.800000	22.599000	72.678200	BU #82 Warren Towers 35
...	...	...	...	...	...	...	...
7161	39880	31186	2022-09-27	502.000000	19.840000	67.712000	BU #69 - Graduate Apartments
7162	39880	31186	2022-09-28	525.333333	17.156667	62.882000	BU #69 - Graduate Apartments
7163	39880	31186	2022-09-29	552.000000	15.150000	59.270000	BU #69 - Graduate Apartments
7164	39880	31186	2022-09-30	590.000000	12.020000	53.636000	BU #69 - Graduate Apartments
7165	39880	31186	2022-10-01	590.000000	14.400000	57.920000	BU #69 - Graduate Apartments

7166 rows × 7 columns

	device_id	serial_no_x	date	valuePsi	celsius	fahrenheit	Site	serial_no_y	Material	Tons
0	39688	31167	2021-08-20	596.000000	30.755000	87.359000	BU #90 - School of Law	31936	Recycling	0.0275
1	39688	31167	2021-08-24	620.000000	27.512500	81.522500	BU #90 - School of Law	31936	Recycling	0.0250
2	39688	31167	2021-08-27	597.714286	34.627143	94.328857	BU #90 - School of Law	31936	Recycling	0.0250
3	39688	31167	2021-08-31	892.444444	22.646667	72.764000	BU #90 - School of Law	31936	Recycling	0.0330
4	39688	31167	2021-09-03	726.400000	19.928000	67.870400	BU #90 - School of Law	31936	Recycling	0.0740
...	...	...	...	...	...	...	...	...	...	...
184	39880	31186	2022-06-07	492.000000	21.960000	71.528000	BU #69 - Graduate Apartments	31894	Recycling	0.0250
185	39880	31186	2022-06-10	684.000000	23.440000	74.192000	BU #69 - Graduate Apartments	31894	Recycling	0.0250
186	39880	31186	2022-06-14	796.000000	26.220000	79.196000	BU #69 - Graduate Apartments	31894	Recycling	0.0300
187	39880	31186	2022-06-16	438.000000	19.980000	67.964000	BU #69 - Graduate Apartments	31894	Trash	3.9700
188	39880	31186	2022-06-24	360.000000	24.230000	75.614000	BU #69 - Graduate Apartments	31894	Recycling	0.0000

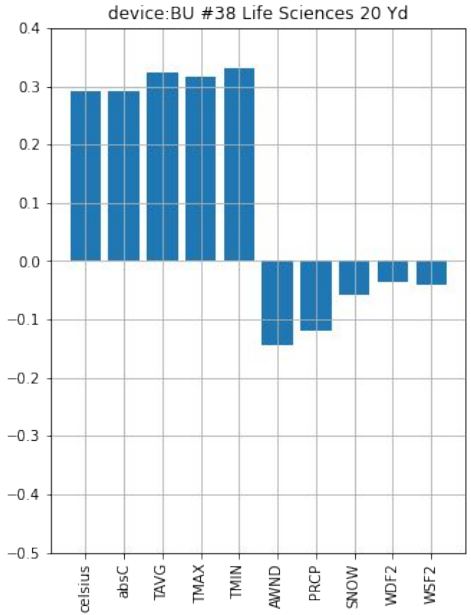
First look by correlation with psi

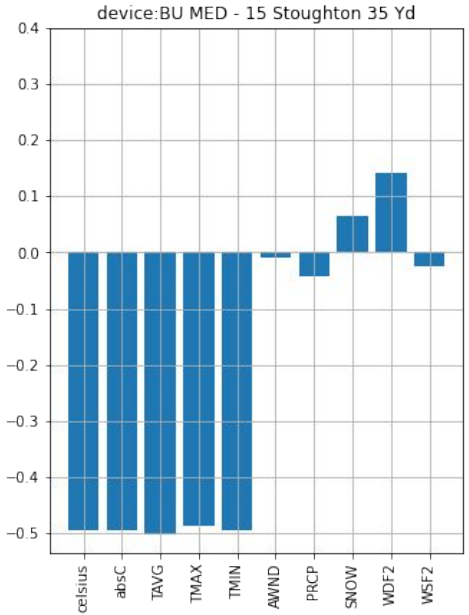
- Correlation is too weak
- We need to divide the data



Further divide by device (2 samples)

- Completely different patterns
- Temperature features work

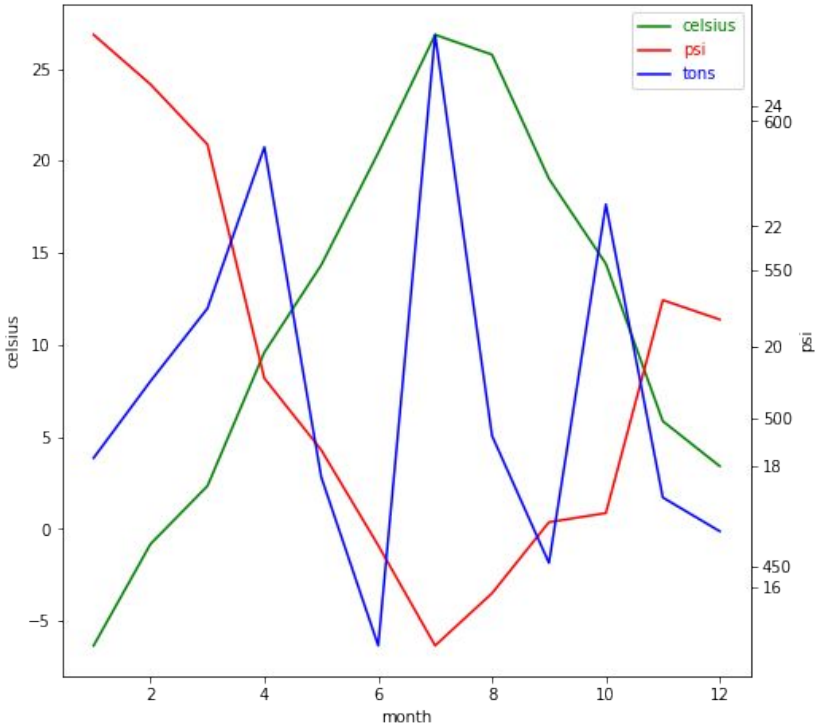




Is psi another form of tonnage?

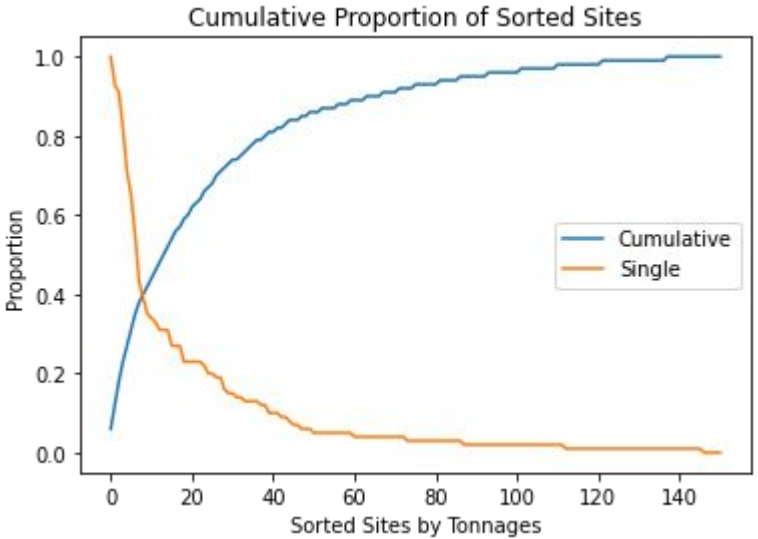
- Basically not
  - Need to find a clever way to utilize psi (future work)
- Why?
  - Cannot sum up psi
  - Averaging psi causes misleading
  - Multiple Psi(s) from short periods(below)

date	psi
2022-01-11T12:15:48	528
2022-01-11T12:16:04	224
2022-01-11T12:18:58	720

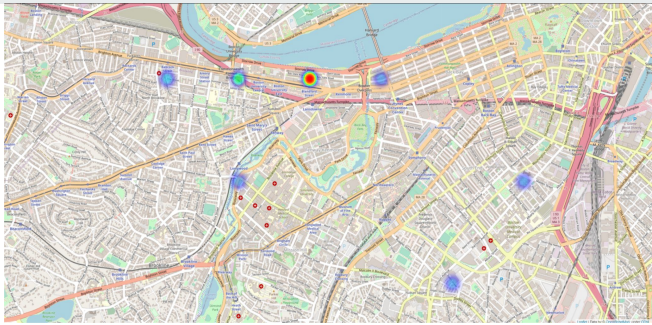


Uneven distribution of waste

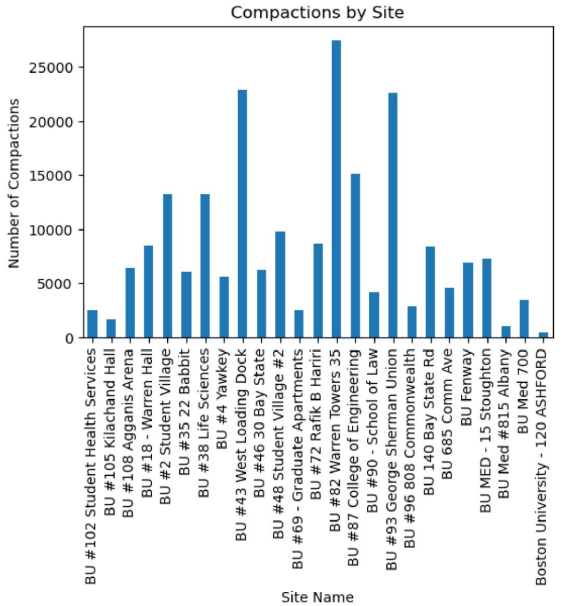
- 40 of 151 sites contributed 80%



Observations

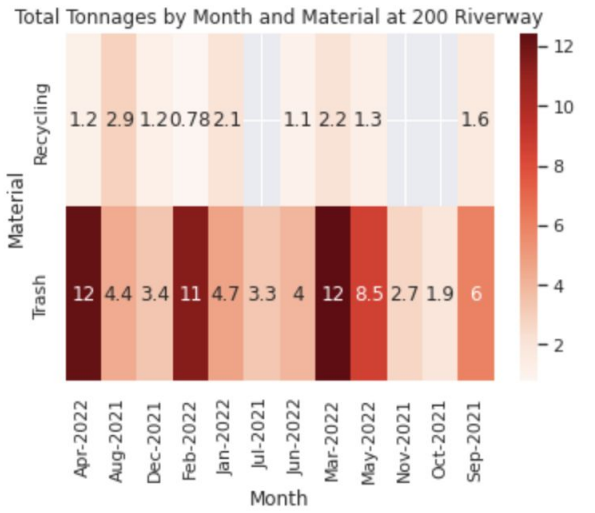


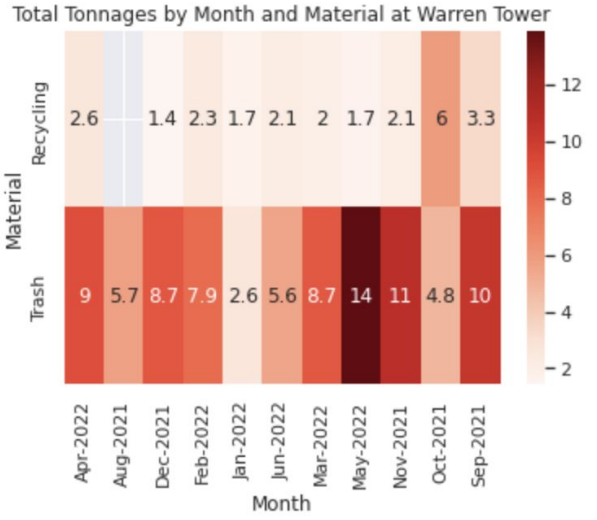
- From compactor location heatmap, we have these observations
  - More devices grouped at main BU campus, less elsewhere
  - No strong correlation between weather and amount of waste has shown
- Some popular locations show more significance than others
  - Warren Tower
  - GSU etc.
  - Will focus more on these sites to get a stronger correlation



Tonnage Changes at Popular Sites

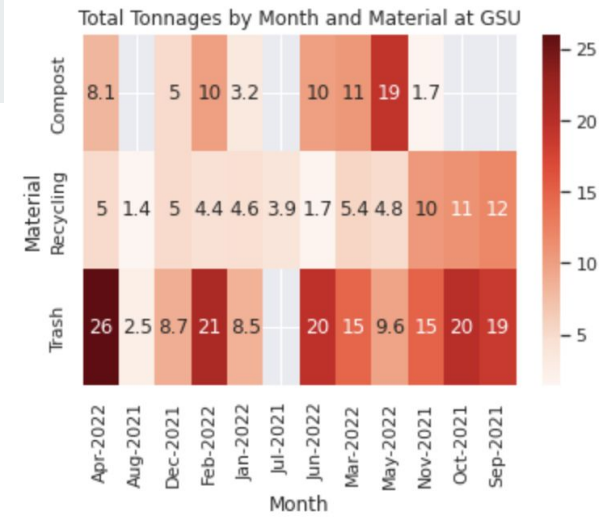
- Selected 200 Riverway, Warren Tower, GSU, and West Loading Dock
- Observations 1: Highest tonnage of trash happens in regular semester
  - At 200 Riverway site, heatmap shows a trend of higher trash tonnage in warmer months
  - However, Warren Tower site doesn't follow this trend
- Observation 2: Less tonnage during summer break
  - Less human traffics, less campus activities

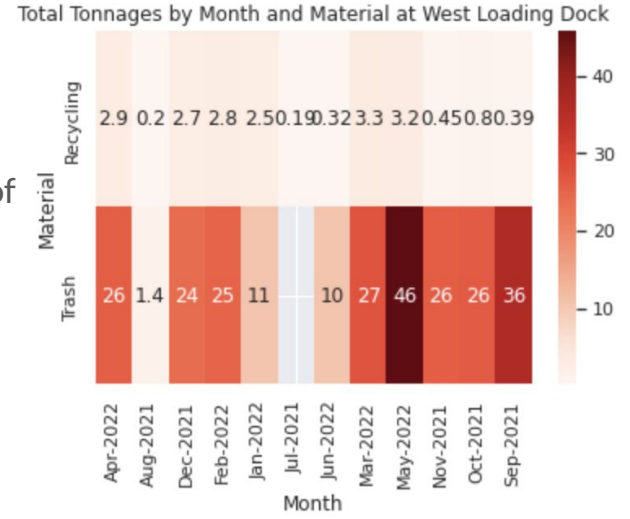




Cont.

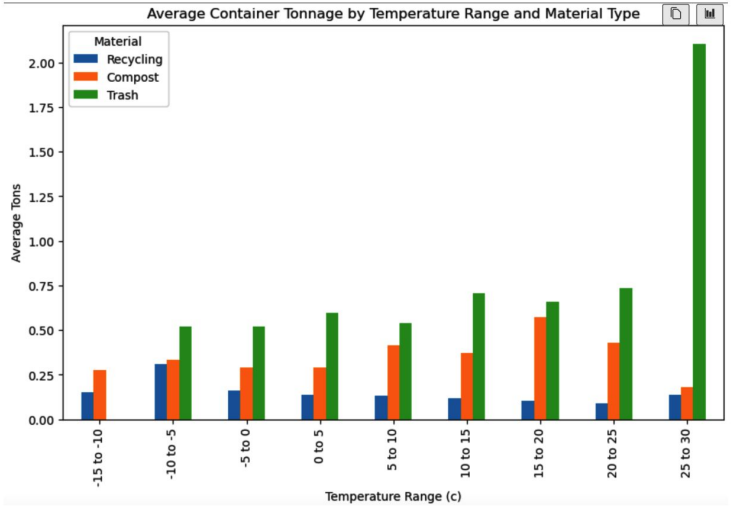
- Heatmap at GSU and West Loading Dock sites
- Observation 1: GSU heatmap shows a huge decrease of trash tonnage during summer break (26 vs. 2.5 tons, April vs. August)
  - Possibly indicates waste has stronger correlation with campus activity
  - Same trend showed in West Loading Dock heatmap
    - (46 vs. 1.4 tons)
- Observation 2: No clear correlation between weather and amount of of waste
  - As indicated in West Loading Dock heatmap
  - A slight correlation shown in GSU heatmap.





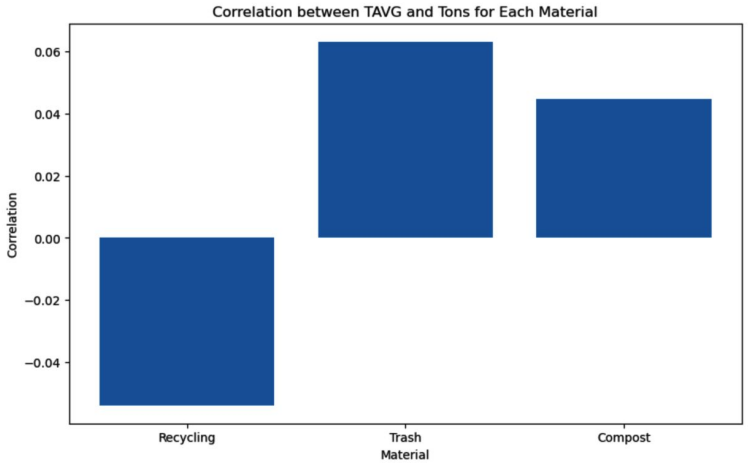
Waste Tonnage & Changing Temperatures

- Process:
  - Combined “Daily Weights” data with external NOAA weather data
  - Assigned a temperature range to each entry
  - Grouped by material type and temperature range, then calculated average of each group
- Results:
  - Signs of correlation?
  - Does waste generation increase with temperature?



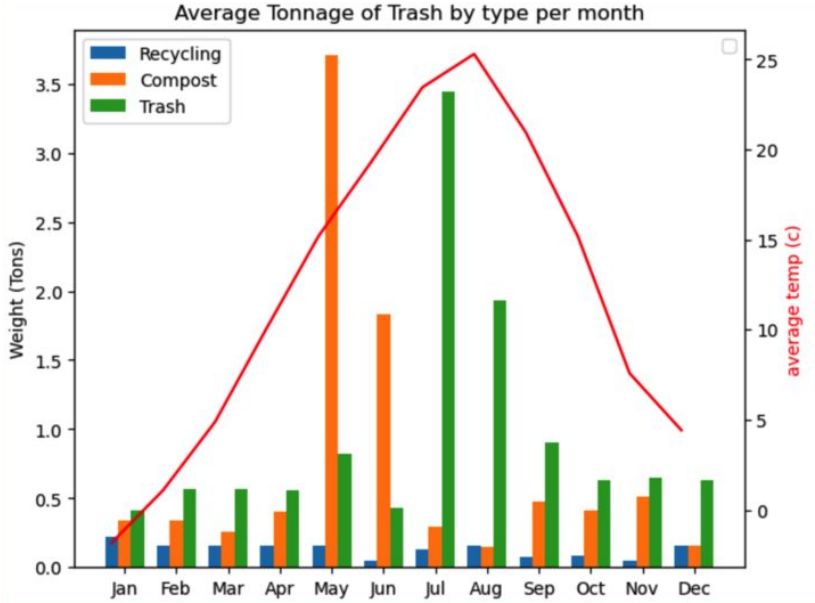
Waste Tonnage & Changing Temperatures

- Process:
  - Calculated the correlation between Daily Avg. Temperatures and Waste by type
- Results:
  - Small increase in Trash and Compost with increasing temperatures
  - Small decrease in Recycling with decreasing temperatures
- Implications?
  - How does waste increase when campus activity decreases?



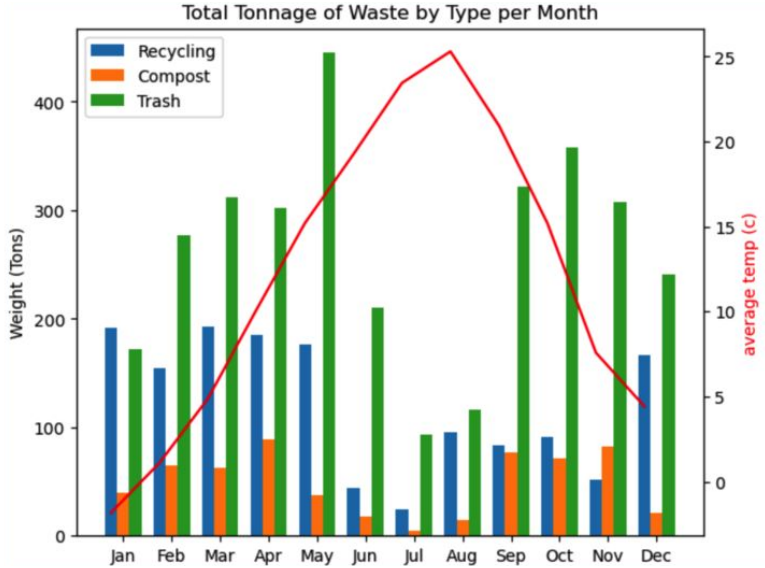
Optimizing the Presentation of the Data

- Process:
  - Grouped by waste type and month
  - Display the average of each group
  - Plot temperature as line chart
- Result:
  - Updated bar chart shows which months of the year showed the highest average waste.
  - Confirming increased activity during the summer months?
  - Still not in line with client's expectations and my assumptions.



Analyzing the data from a different perspective

- Process:
  - Grouped by waste type and month
  - Display the **total** waste of each type per month
  - Plot temperature as line chart
- Result:
  - Updated bar chart shows expected results
  - Decreased total waste generated in the summer months, increased in the fall / spring semesters.
  - Peak waste generated in May due to Move Out & Commencement
- Implications?
  - Average monthly waste may be inversely correlated with total monthly waste
  - Increase waste generation across campus, decreased waste generation overall / at the popular sites



Challenges, Limitations and Assumptions

- We have also encountered some challenges with the data
  - Some of them are too noisy/irrelevant
  - Noise from people's daily life
  - Difficulty extracting and combining the relevant documents provided
    - Mismatching addresses for compactors
- Limitation of the term "Weather"
  - Many factors can affect the weather, not only temperature,
    - Makes it hard to tell a direct correlation
- Assumptions
  - Patterns are not changed during the period of data collection

Future Work

- Combine all the data provided to us in order to gain more accurate results
    - Allows us to form better connections between different aspects of the data
    - Centralizes the data to make it more easily accessible to all team members
- Denoise the data by removing container sites that are do not see much activity
- Recalculate correlation coefficients between waste generation and weather
    - Explore how average daily temperatures affect the most active container sites
- Explore additional drivers of waste generation such as campus activity
    - Early observations show more waste generation during warmer weather, when campus activity is expected to be lower, this will need to be examined for accuracy