Deliverable 1 Report

Project: Sidewalks
Team: 1

Members:

1) Quang Nguyen
2) Anargh Sidhardhan
3) Akarvin Raja Prajan
4) Pranesh Jayasundar

Deliverable Checklist : **Status- Completed**

- Collect and pre-process a preliminary batch of data
- Perform a preliminary analysis of the data
- Answer 1-2 key questions
- Submit all of the following information (code, notebooks, answers to questions) as a PR to your team's branch on github. (Add your PM and TE as reviewers!)
- Submit the Weekly Scrum report to the gradescope and upload to google drive.
    - Make sure to identify which team member is doing which tasks on the scrum report

Analysis:

The team has conducted an initial analysis of the dataset "Boston Sidewalks Data".

Analysis on Ramp_0 Dataset:

- If a ramp 'x' is located near other ramps that are in excellent condition, then that particular ramp would have a better score. This can be done by checking the feature - 'INEARFID', and get the 'RAMP_COND' of that particular ramp id - FID.
- The significance of the features 'CREEC_Type', 'CREEC_Posi' is unknown to draw conclusions from or form relation with other features of the dataset.
- For columns: 'DWP_COND', 'LANDING_CO' discrepancies were observed. For instance, for a particular ramp 'y,' the value of one column is Excellent, whereas the other column says 'Fair'. It is nondeterministic to identify the weightage of features that will determine the ramp condition. These columns share the same values but there are 12159 rows for which these two columns value mismatch.

    - Excellent (Like new)
    - Poor (Panel is damaged and affecting functionality- to be replaced)
    - Fair (panel shows wear but is still fully functional) [or] Fair (Minor/Hairline damage with no impact to accessibility)
- Columns: 'Failure_Ty', 'Failure_La' - These two columns helped in identifying the primary reason for poor sidewalks and excellent sidewalks.
    - Example:
        - Concrete FailureDWP FailurePvt/Cswk Failure
        - Concrete FailurePvt/Cswk Failure
        - Poor Ramp Condition
        - Poor Ramp ConditionCrosswalk missing/misaligned
- Columns: 'NEAR_FID', 'NEAR_DIST' - These columns have a value of -1, which doesn't provide much information.
- Columns: 'I_NEARFID' - With this column, it is possible to identify the adjacent/nearest ramp condition to every other ramp.
- Columns: 'Crosswalk_', 'RAMP_COND' - There is a difference between these 2 columns, where the values differ. It is unclear how the ramp and the crosswalks are related in the same coordinate points.
    - Poor- needs to be re-striped
    - Fair- slightly faded
    - Excellent- like new condition
    - Excellent (Like new)
    - Fair (Minor/Hairline damage with no impact to accessibility)
    - Poor (Panel is damaged and affecting functionality- to be replaced)
- Columns: 'RAMP_MATL' - Most of the ramps are made of Portland Cement Concrete [excellent, fair, poor].

Additionally, using the latitude and longitude, postal code and address was extracted.

In order to visualize this dataset, on the basis of the ramp condition using the folium library, the co-ordinates were plotted on the map with respect to the excellent, fair and poor conditions.

Below tables, identifies the top 5 locations based on the postal and neighborhood to discern the locations where the ramps are in good condition, poor condition, and fair condition:

Top 5 neighborhoods where the Ramps that are in excellent condition

Neighborhood	Postal	Count
West Roxbury	02132	891
South Boston	02127	888
Roxbury	02119	816
Roslindale	02131	679
Dorchester	02124	659

Top 5 neighborhoods where the Ramps that are in fair condition

Neighborhood	Postal	Count
Mattapan	02126	414
Dorchester	02124	389
East Boston	02128	345
South Boston	02127	270
Charlestown	02129	269

Top 5 neighborhoods where the Ramps that are in poor condition

Neighborhood	Postal	Count
Roslindale	02131	50
Roxbury	02119	43
Dorchester	02122	42
Hyder Park	02136	42
South Boston	02127	37

Analysis of the Sidewalks_Hazards Dataset:

The sidewalks hazards are classified into three types:

1) Fixed pinch point <36” sidewalk width
2) Trip hazards due to tree roots
3) Trip hazards not due to tree roots

Most of the hazards are of type trip hazards not due to tree roots.

In addition to the existing dataset, using the latitude and longitude coordinates, the address of the hazards were extracted. This can help identify neighborhoods with most hazards that would require immediate attention for the safety of the pedestrians.

The top three neighborhoods with these hazards are:

Hazard Type/ Neighborhood	Fixed pinch point <36” sidewalk width	Trip hazards due to tree roots	Trip hazards not due to tree roots	**Total hazards**
1) Highland	56	220	227	**503**
2) Dorchester Heights	59	52	149	**260**
3) Mount Hope	8	61	75	**144**

Using the coordinate system, and the folium library, the different types of hazards were plotted on the map. For each sidewalk hazard, there is a map generated of 10 cluster regions to identify the regions that have more concentrated hazards that need attention.

Analysis of the Sidewalks Dataset:

From this dataset, it was possible to identify the sidewalks in a region and their characteristics such as width and area.

The dimensions of the sidewalks are assumed to be given in feet from the sidewalks design documentation and older sidewalks’ metadata.

Top 5 Districts with the most number of sidewalks:

District Name	Number Of Sidewalks
South Dorchester	3017
North Dorchester	266
West Roxbury	2537
Downtown	2055
Allston/Brighton	2041

Top 5 districts with the largest sidewalks' width:

District Name	Sidewalk Width(ft)
East Boston	8.81
Downtown	8.64
Roxbury	8.33
Charlestown	8.14
North End	7.89

Top 5 districts with the largest sidewalks' area:

District Name	Sidewalk Area (sq ft)
Roxbury	4670.44
Roxbury 10A	3046.55
East Boston	2908.68
Downtown	2864.80
The North End	2682.57

The dataset from Boston's public schools provided information on the number of sidewalks that are accessible within a 15-minute walking distance (1.2 kilometers approx.)

Top 5 Schools with most sidewalks within 15 min walk.:

School Name	Number Of Sidewalks
Eliot K-8	812
Kenny Elementary	492
Quincy Lower (K-5)	469
Lnydon K-8	463
Kilmer Upper (4-8)	458

Analysis of the Roadway Centerline Dataset:

This dataset contains road centerline data in Boston. A roadway centerline is a vector line data that represents the geographic center of a roadway between road shoulders. It can be used to encode road characteristics, such as road name, type, speed limit, etc. It can also be used for linear referencing systems (LRS), which allow locating features along a route using distance measures. This dataset shows which roads are public or private, and what their functions are. It also shows which districts they cross. This could help with mapping. To get accessibility scores for each road, we need more data than this dataset, but some of its features might be helpful with other datasets.

Number of Roads per jurisdiction (Top 5):

Jurisdiction	Number of Roads
Public way	3528
Private way	1215
Public way/Private Way	222

DCR	47
Town Line	13

Number of Roads per District (Top 5):

District	Number of Roads
3-07	599
2-06	569
3-03	506
2-08	452
2-04	447

Number of Roads per Council (Top 5):

Council	Number of Roads
6	803
5	710
1	676
7	585
2	543

Limitations:

The description document from the client is pending. Without this, it is difficult to understand the significance of the features in the dataset to extract the accessibility score.

Key Questions:

1. How can we leverage existing data (trip hazards, pedestrian ramps, crosswalks, etc) to create an accessibility score for each sidewalk block in the City? Can we use this score to identify which sidewalks are the most inaccessible in the City?

- We believe this is feasible, we can look for the ramps that are in good condition, check for its adjacent ramps/crosswalks, look for any trip hazards if present and create an accessibility score accordingly. Additionally we can check for features that support the healthy condition of the ramp.

2. Can we use the results of this score to identify regions of the City that are the most accessible vs. least accessible? Can we also compare these results with different elements of social vulnerability to see how equitable/inequitable accessibility is?

- Yes, we can use the accessibility score to identify regions in the city that are most accessible vs. the least accessible. Currently, we are able to observe the excellent ramps, fair ramps, and poor ramps around Boston, and we listed the top 5 neighborhoods and their postal code. Going forward we should be able to do the same after generating the accessibility score.