APPLIED DATA SCIENCE CAPSTONE

April 8, 2020

HOW FAR

FROM POINT ‘A’

TO POINT ‘B’?

INTRODUCTION

Having access to accurate distance information (or distances, if there are multiple locations) relative to a known object or location can aide in decision making.

For example, in the event of a medical emergency while in an unfamiliar city, it would be useful to know how many hospitals are nearby and which one is the closest.

INTRODUCTION CON'T.

This project will create a process for calculating distance from a specific location to a set of facilities. The process will be illustrated using an example based in Seattle, Washington, which is a densely populated urban area in the northwestern United States.

The specific location chosen for this illustration is Seattle's Space Needle. It is a city landmark and is considered an icon of Seattle. It was built in 1962 for the World's Fair and stands 605 feet tall.

In addition, since the number of beds varies across hospitals, an attempt will be made to see the extent to which the hospitals returned by the API search will cluster based on number of beds.

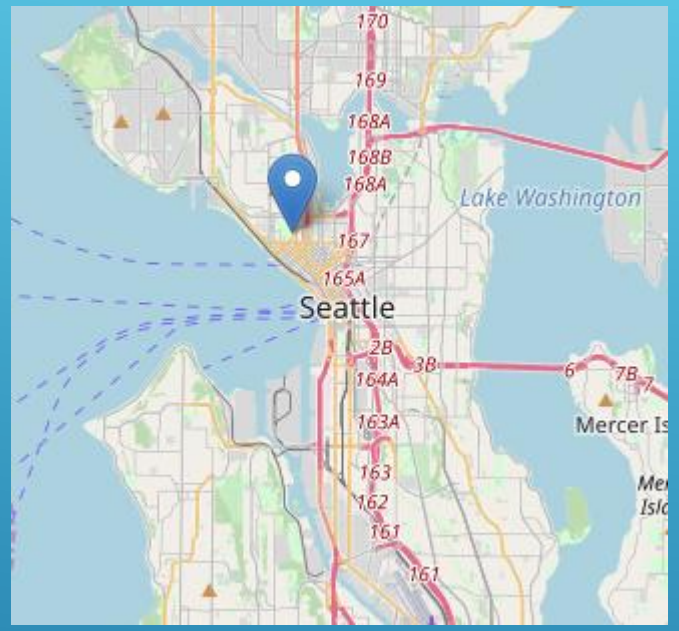
DATA

The Foursquare API will be used to obtain location data for acute care hospitals that are within 20 miles from the Space Needle.

In addition, hospital bed information will be obtained from two sources:

https://en.wikipedia.org/wiki/List_of_hospitals_in_Washington_(state)

https://www.wsha.org/our-members/member-listing/



The Nominatim package from GeoPy was used to obtain the latitude and longitude for the Space Needle as well as the geographic boundaries of Seattle.

The search results obtained from Foursquare were transformed into a data frame consisting of 50 rows.

METHODOLOGY

METHODOLOGY CON'T.

Further review revealed the following data issues:

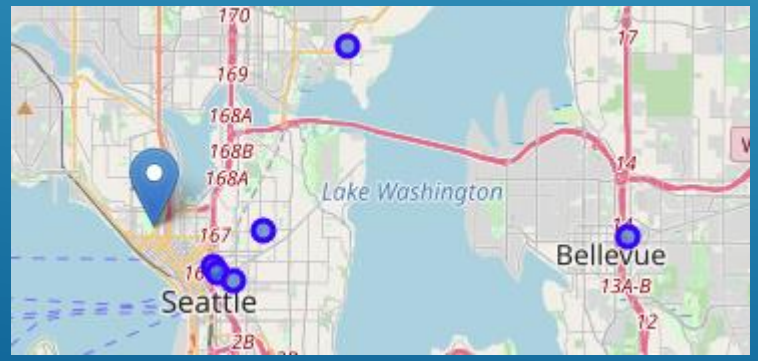
- Many non-acute hospital facilities were returned by the Foursquare search
- NaN values for address
- Duplicate addresses
- Hospitals that had closed

Based on this review, rows were dropped for all the above conditions, resulting in a new data frame consisting of six rows. The straight-line distance from the Space Needle to each hospital was calculated using the geodesic function.

RESULTS

Based on the selection criteria, six acute care hospitals were identified within a twenty-mile radius of the Space Needle, as displayed below.

	name	categories	location.address	location.city	location.lat	location.lng	distance	beds
0	Kindred Hospital Seattle - First Hill	Hospital	1334 Terry Ave	Seattle	47.611843	-122.328792	1.129980	50
1	Virginia Mason Hospital and Seattle Medical Ce...	Hospital	1100 9th Ave	Seattle	47.610128	-122.327232	1.256082	336
2	Swedish Medical Center - First Hill Campus	Hospital	747 Broadway	Seattle	47.608403	-122.321890	1.529688	697
3	Kaiser Permanente Hospital #1	Hospital	201 16th Ave E	Seattle	47.620243	-122.311679	1.757514	18
4	Seattle Children's Hospital	Hospital	4800 Sand Point Way NE	Seattle	47.662590	-122.282741	4.255469	407
5	Kaiser Permanente Hospital #2	Hospital	11511 NE 10th St	Bellevue	47.618550	-122.186462	7.607533	347



RESULTS CON'T.

The second objective of this project was to determine whether hospitals would cluster based on number of beds. Since the analysis data set is so small, visual examination suggests that hospitals appear to cluster into three groups, which was confirmed by conducting a k-means analysis.

Hospital	Beds	Manually Determined Cluster	K-Means Cluster
Kindred	50	Small	1
Virginia Mason	336	Medium	0
Swedish	697	Large	2
Kaiser #1	18	Small	1
Children's	407	Medium	0
Kaiser #2	347	Medium	0

CONCLUSION

The two objectives of this project were completed, with the following results:

- Six acute hospitals were identified within a twenty-mile radius of the Space Needle.
- The six hospitals clustered into three groups based on bed count.