Lake Michigan Influences

Blake Wallace
Capstone Technical Report

May 17, 2019

Data science objectives:

1. Does Rain have any effect on the average daily temperatures?
2. What effect does rain have on the average daily temperature near the water?
3. What effect does rain have on the average daily temperature far from the water?
4. When it rains, is there a statistically significant difference between the amount of rain that falls in downtown Chicago compared to the Ohare airport?
5. How much correlation exists between the average daily temperature of Lake Michigan and the temperature difference between the downtown Chicago area and the Ohare airport?
6. Can we build a model that predicts with at least 80% accuracy the difference in total precipitation between Ohare airport and the the Botanical gardens?
7. Is there a statistically significant difference between the daily temperature near the water as apposed to far from the water?

Data Sources:

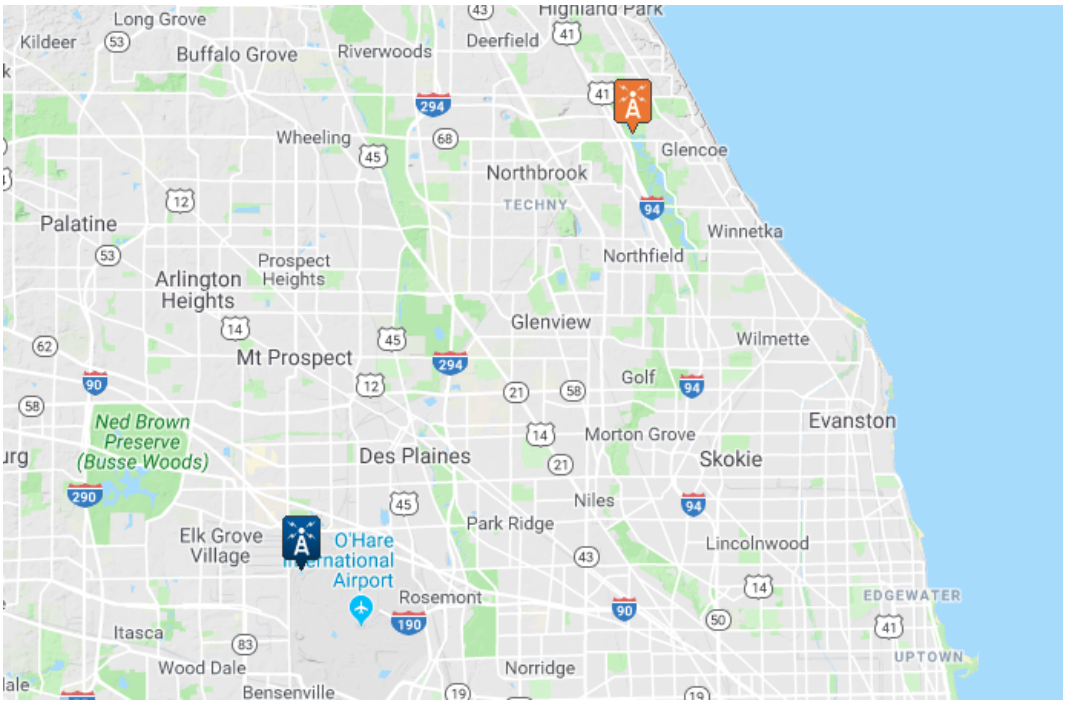

Figure 1: In the top right is the location of the weather tower inside of the Chicago Botanical Gardens, while the bottom left shows the location of the tower in the O'Hare Airport.

CHICAGO OHARE INTERNATIONAL AIRPORT, IL US
- Source: National Centers for Environmental Information
- GHCN (Global Historical Climatology Network) Daily Documentation
- ID GHCND:USW00094846
- 41.995 N 87.9336 W
- Airport Information

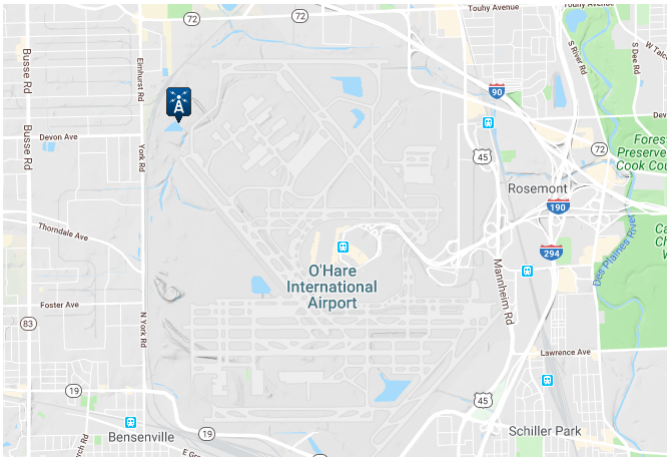

Figure 2: In the top left is the location of the weather tower inside of O'Hare Airport.

CHICAGO BOTANIC GARDEN, IL US
- Source: National Centers for Environmental Information
- GHCN (Global Historical Climatology Network) ? Daily Documentation
- ID GHCND:USC00111497
- 42.13987 N 87.78537 W
- Garden Information

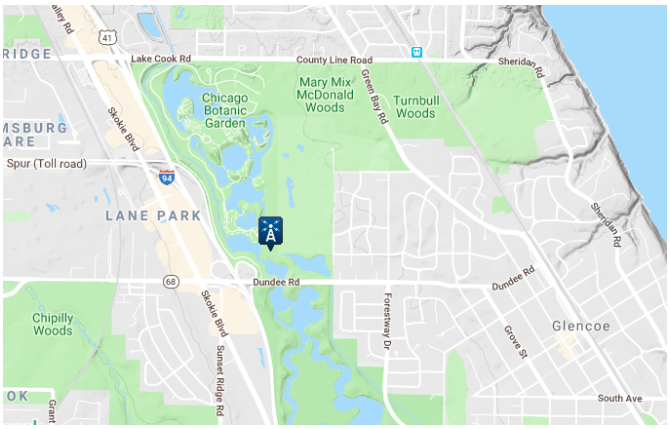

Figure 3: The weather tower at the Chicago Botanical Gardens.

Lake Michigan
- Source: Great Lakes Statistics: Average Surface Water Temperature from the Great Lakes Surface Environmental Analysis (GLSEA)
- 44.0 -87.0 (44 00' 0.00" N 87 00' 0.00" W)
- Data Set for 2018

Station FSTI2 - Foster Ave., Chicago, IL
- Source: National Data Buoy Center
- Owned and maintained by Chicago Park District
- 41.976 N 87.648 W (4158'35" N 8738'51" W)

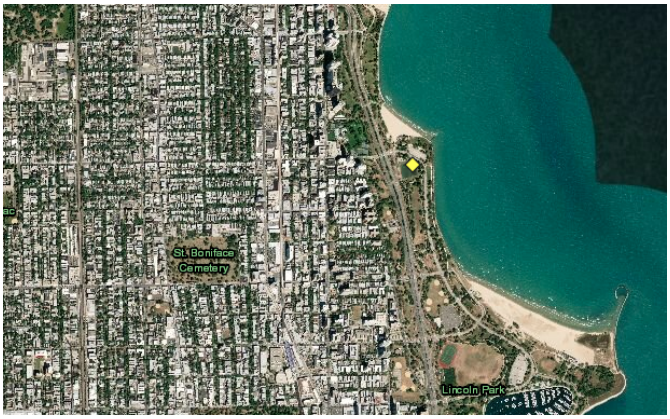

Figure 4: In the top left is the location of the weather tower inside of O'Hare Airport.

Data

Links

Data Dictionary
Bouy Data Dictionary
O'Hare Airport Data Dictionary
Botanical Garden Data Dictionary

"The five core values are:"

ohare_prcp - Precipitation (PRCP) (inches)
ohare_snfall - Snowfall (SNOW) (inches)
ohare_sndpth - Snow depth (SNWD) (inches)
ohare_maxtmp - Maximum temperature (TMAX) (Fahrenheit)
ohare_mintmp - Minimum temperature (TMIN) (Fahrenheit)

Other Features

lake-temp - Average Daily Surface Water Temperature for Lake Michigan (Fahrenheit)
garden_prcp - Precipitation (PRCP) (inches)
garden_maxtmp - Maximum temperature (TMAX) (Fahrenheit)
garden_mintmp - Minimum temperature (TMIN) (Fahrenheit)
garden_tobs - Temperature at time of observation (TOBS) (Fahrenheit)
ohare_wspd - Average daily wind speed (AWND) (miles per hour)
ohare_atmp - Average Temperature (TAVG) (Fahrenheit)
ohare_w2dir - Direction of fastest 2-minute wind (WDF2) (the direction the wind is coming from in degrees clockwise from true N)
ohare_w2spd - Fastest 2-minute wind speed (WSF2) (miles per hour)

Feature Engineering

target - absolute difference between the precipitation measurements at Ohare and the garden (ohare_prcp - garden_prcp)
garden_didrain - categorical, 1 for yes, 0 for no
ohare_didrain - categorical, 1 for yes, 0 for no
garden_medtmp - Median daily temperature at the Garden/ midpoint between the max and min temperatures ((garden_maxtmp + garden_mintmp)/2)
ohare_medtmp - Median daily temperature at ohare/ midpoint between the max and min temperatures ((ohare_maxtmp + ohare_mintmp)/2)
tmpdiff - difference between the median temperatures at ohare and the garden (ohare_medtmp - garden_medtmp)

Data Cleaning/Data Manipulation/EDA:

Figure 2 shows the .

Tests and Evaluation:

Table 1: Statistical Tests with Results

Data	t-score	p-value	Significance	Gardens Avg (F)	Ohare Avg (F)
All Data	0.5876	0.5568	None	59.24	59.43
No Rain	3.285	0.0010	Yes	58.99	60.57
Both Rain	-2.629	0.0086	Yes	59.48	57.7
ohareRain	-1.9557	0.0506	None	59.06	57.43
gardensRain	0.0904	0.9280	None	59.99	60.07

Models and Evaluation:

Table 2: Predictive Models with their scores

Model	Training score*	Testing score*	Training MSE**	Testing MSE**	Cross Validation
Linear no poly	0.0825	0.1052	0.0933	0.0683	0.0785
Linear gs	0.1222	0.1329	0.0893	0.0662	0.0984
Decision Tree	0.1139	0.0691	0.0901	0.0711	0.0429
Decision Tree gs	0.0937	0.0584	0.0922	0.0719	0.0450
Random Forest	0.8614	0.0517	0.0134	0.0724	0.0554
Random Forest	0.8711	0.1078	0.0131	0.0681	0.0770
Random Forest gs	0.8658	0.0905	0.0136	0.0694	0.0651
Random Forest	0.8677	0.0682	0.0135	0.0711	0.0660
Random Forest	0.8704	0.1153	0.0132	0.0676	0.0767
Random Forest	0.8080	0.0957	0.0195	0.0690	0.0787
Random Forest ada	0.9547	0.0549	0.0331	0.0722	0.0331
Random Forest ada	0.9445	0.0525	0.0056	0.0723	0.0283
Random Forest bag	0.6735	0.1130	0.0332	0.0677	0.0928
Random Forest bag	0.6705	0.1239	0.0335	0.0669	0.0943

* The score refers to the Coefficient of Determination.
** MSE - Mean Squared Error
gs denotes a Grid Search was performed.
ada denotes an Ada Boost model was performed.

Resources: