Michigan Outdoor Recreation Search Interest Installment 10

This is my project to analyze and forecast Google search interest for 10 forms of oudoor recreation by people in Michigan. My data is daily data for each form of search interest from January 2021 to March 2025. The search interest terms for this project are atving, boating, camping, fishing, hiking, kayaking, rving, hunting, skiing, and snowmobiling.

So far, I have attempted three different models to forecast this data. I have chosen my preferred model for each of the 10 forms of outdoor recreation.

However, throughout these installments, I have used actual values for weather data which is a key driver for how the residual variation of search interest differs from seasonal expectations and long term trends. These weather variables used in the model include data up to the day of the search interest. Naturally, if you are forecasting search interest into the future dependent on data also in the future then you must also have forecasts for these independent variables.

This is a common 'problem' when forecasting based on exogenous regressors. Generally, these regressors are acquired from a different source and adjusted based on the task at hand.

In previous installments, I have used weather variables for maximum daily temperature, minumum daily temperature, precipitation, and snow depth. These have been the average of 105 weather stations in Michigan which meet criteria related to data availability. Please see the below link for more information about how this data was pulled and aggregated.

https://dataandoutdoors.com/michigan-outdoor-recreation-search-interest-installment-iii/

However, I never wanted to pull 105 weather stations of data when engaged in weekly forecasting. But I didn't want my results to be based on this choice. Generally, it's far easier to adjust how I source the weather data than the model itself. In order to implement the models for forecasts, I will only pull data from five places in Michigan: Detroit, Grand Rapids, Houghton Lake, Sault Saint Marie, and Marquette. These are chosen based on their geographic location and proximity to population centers.

This is possible because weather is highly correlated accross space. It's not necessary to account for the average of 105 places because different plades similarly located will have the same weather on the same day. One option for reducing the places is by looking at statistical criteria for which places explain the most variation of the aggregate. However, in my experience, when the possible explanatory variables are highly correlated (as with weather in nearby places) different variables are similar in efficacy and the best variable is chosen due to random variation and does not make the best intuitive sense.

Below are the R Squared when regressing the average of the 105 places used in model development on the five places I've chosen for implementation. In each case, the weather in the five places explains most (at least 2/3) of the variation in the final average. Precipitation is the lowest, by far. This is most likely because daily rain fall varies the most across space. It is also because the average precipitation for 105 places is rarely zero because on most days is rains at least some in at least one place. Conversely, on most days an inidividual place has no rain. It's difficult to explain these near zero average rain falls in the 105 place average using a handful of places. I attempted adding up to another five places to the regression without considerable improvment. Thankfully, the fact that it rains in a small fraction of places is unlikely to have a large influence on statewide search interest.

The other three weather variables have an R-Squared of at least 90%. Therefore, when creating forecasts I will only pull data for these five places. I will not simply average the values of these five places, but rather use the estimated OLS regression to translate these values to the 105 place average used in model development.

Variable	R Squared
Maximum Temperature	99%
Minimum Temperature	99%
Precipitation	68%
Snow Depth	92%

Finding external forecasts for maximum and minimum temperature and precipitation is relatively easy. However, forecasts for snow depth are not as common. Thankfully, for this project I'm only forecasting two weeks ahead which means that the current snow depth is fairly informative of the future. What I want to determine is how the current snow depth is likely to change over the next two weeks from each forecast start point. The current snow depth could increase due to further snow fall or decrease due to melting.

While it's difficult getting future forecasts for snow depth, getting future forecasts for snowfall is easier. I assume that snow fall directly increases snow depth. So if there is one inch of snow fall, the snow depth at the given location increases by one inch.

For snow melt, I look at my entire weather dataset for my five places. For each place I look at only days where the previous day there was positive snow depth. I then restrict my data to only days where there is no precipitation (or potential positive snowfall). Then I regress, without intercept, the change in snow depth on the number of degrees that the maximum temperature exceeds 32 (the freezing point). The chart below shows the number of inches of snow melt expected for each degree the maximum daily temperature exceeds freezing. Each place has a similar value ranging from 0.07 of daily snow melt per degree over freezing to 0.12.

Place	Snow Melt
Detroit	-0.07
Grand Rapids	-0.12
Marquette	-0.08
Houghton Lake	-0.07
Sault Saint Marie	-0.08

To get the snow depth for each day in the forecast, I take the original snow depth and apply the amount of snowfall as a positive change in snow depth whenever there is forecasted snowfall. I also apply the amount of predicted snow melt for each day where the maximum temperature exceeds 32 degrees. Then I limit the snow depth at 0 inches minimum. Then I unwind to the next day and the next for the fourteen day forecast.

Finally, I would like a convenient method for getting the forecasts for maxium temperature, minimum temperature, precipitation, and snow fall. For my Northern Michigan Search interest project, I manually look up this information on the Accuweather and weather.gov websites. That's because the Northern Michigan project uses three month forecasts which are generally only available on expensive subscriptions for weather apis. However, since my forecasts are only two weeks ahead for this project, I use the open weather map API. I choose this API over accuweather to avoid an expensive subscription even for limited data. The benefit of the API is that I can automatically access the website data using the requests python package.

The open weather API provides daily weather forecasts for one week on their free/near free weather api. Then after the first week, open weather provides long term forecasts. I'm not entirely sure what the difference is between the daily and long-term forecasts other than the api call is different. However, I believe the daily forecast uses a more sophisticated/accurate forecast methodology. The long-term forecast uses historical averages and, it appears, long-term climate predictions. Note that if you pay for a subscription each month you could get the daily forecast for a longer period of time. However, for my purpose I will use the free version. The link below provides the documentation for the open weather API.

https://openweathermap.org/api/one-call-3

One issue with the long-term api is that it doesn't provide forecasts for snowfall but only total precipitation in liquid units. Therefore, I assume that precipitation is snowfall when the maximum temperature for the day is 32 or under. While it's possible to have snow on a day above freezing, often this snow won't accumulate anyways. I multiply the liquid precipitation by 10, because rain is 10 times as dense as snow.

Below is an example forecast I took for Houghton Lake starting on October 25th and 14 days forward.

	Date	minimum	maximum	rain	snow	precipitation
0	2025-10-25	32.25	50.77	0.000000	0.0	0.000000
1	2025-10-26	34.84	52.93	0.000000	0.0	0.000000
2	2025-10-27	35.55	51.48	0.000000	0.0	0.000000
3	2025-10-28	36.18	52.45	0.000000	0.0	0.000000
4	2025-10-29	37.33	48.13	0.000000	0.0	0.000000
5	2025-10-30	33.78	50.41	0.000000	0.0	0.000000
6	2025-10-31	33.73	49.98	0.000000	0.0	0.000000
7	2025-11-01	36.99	47.95	0.005906	0.0	0.005906
8	2025-11-02	40.95	43.24	1.154724	0.0	1.154724
9	2025-11-03	39.39	45.66	0.176378	0.0	0.176378
10	2025-11-04	35.19	49.29	0.000000	0.0	0.000000
11	2025-11-05	39.61	48.44	0.307480	0.0	0.307480
12	2025-11-06	34.44	42.95	0.017323	0.0	0.017323
13	2025-11-07	32.98	45.84	0.005906	0.0	0.005906
14	2025-11-08	33.06	40.47	2.107480	0.0	2.107480

In summary, sourcing forecasts for exogenous covariates is a common problem when forecasting time series data. Often, these forecasts are sourced from elsewhere, and here I choose open weather map api. However, it's also commom that you can't get exactly get exactly the data you want, which leads you to get creative by transforming what is available.

In actual production the process will be the opposite order to what appears in this narrative. First, the open weather map API will be used to pull maximum temperature, minimum temperature, precipitation, and rainfall forecasts for Detroit, Grand Rapids, Houghton Lake, Sault Saint Marie, and Marquette. Then, the current snow depth, forecasted snowfall, and forecasted maximum temperature will be used to forecast the snow depth. Finally, these weather forecasts for these five places will be translated into 105 place average forecast using OLS linear regression.

This is the last main installment for this project. Before the end of the year, I will write an easy to read summary of the basic results and findings of this project. During the winter 2026, I will implement these models to provide frequent forecasts of outdoor recreation search interest in Michigan.