Michigan Outoor Recreation Search Interest: Fifth Installment

==========

This is a refresh of the first installment with updated data to March 2025 (previous end date was April 2024). I will not redo installments 2-4. Instead, I will proceed to employ around three different models/forecasting methodologies for the data with the intent of providing two week ahead forecasts. From these options, I will choose one.

==========

There are many outdoor recreation activities popular in Michigan. Many people websearch these activities on Google and other search engines. The goal of this project is to analyze and forecast the trends of Google Search interest within Michigan for 10 different forms of outdoor recreation: atving, boating, camping, fishing, hiking, kayaking, rving, hunting, skiing, and snowmobiling. Search trends for these activities are likely related to the popularity and interest in these activities in general.

In these project, I use Google trends/pytrends to pull daily data from Jan 2021 to March 2025. For this first installment, I will discuss steps required to pull and clean data, show trends for all ten activities, and investigate correlation between them.

==========

Data

Data is pulled from Google trends using the pytrends python package for each search term from Michigan only. All data trends are normalized between 0 and 100 representing the Google search interest for the search term on that day relative to other days in the data pull. The highest period value for each search term is 100.

For each of the outdoor activities, both the base and present participal form are used, i.e camp and camping. (Depending on the activity, one or the other is generally more common.) For each activity, both word forms are added together and renormalized between 0 and 100.

Due to limitations when pulling data with pytrends, daily data is only available when pulling data less than 6 months at a time. Since each one of these data pulls is internally normalized from 0 to 100, different data pulls are not directly comparable. Therefore, I pull many six month intervals that overlap by three months. This three month overlap allows me to combine two series and re-normalize them to the same scale. Whichever series is lower during the overlapping period is the series that has the highest value. Therefore, the average ratio between the two series during the overlapping period is used to adjust this series downward. This puts both series on the same level. (In the previous versions, I rescale the data series by

standard deviation but I no longer do this.) Once two time periods are combined, the combination is iteratively combined with the other periods.

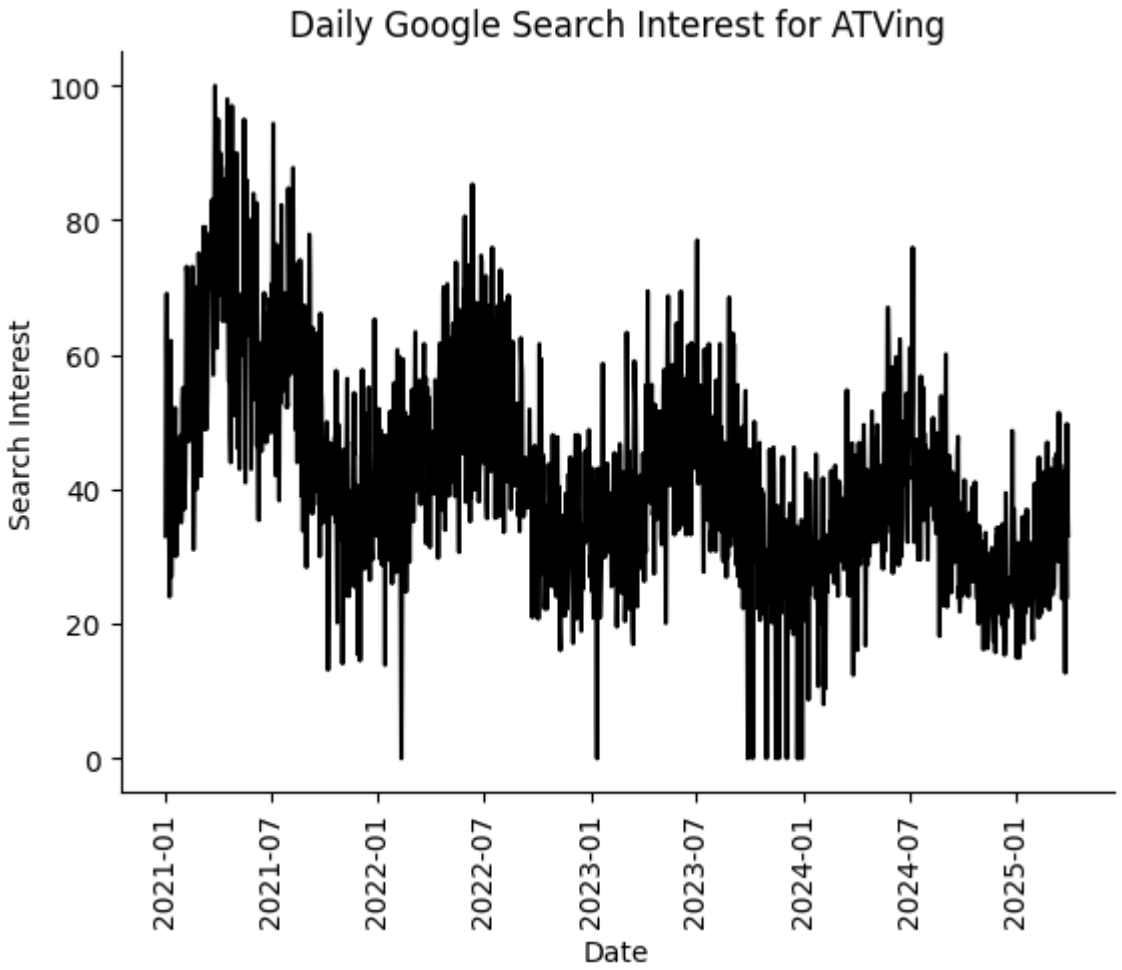
Finally, there are many extremely high search interest values that I identify as outliers. These outliers aren't completely at random in that they appear more likely to occur during times of high search interest. However, these are single day values that are several times higher than adjacent days. I determined that these values are either measurement error or, in the least, values I'm unlikely to explain using forecasting and analysis. Therefore, I remove these by setting values after large one period changes to missing and linearly interpolating the values using the adjacent days' search interest values.

While the resulting dataset is patched together and imperfect, it is what we have and I believe the data will contain a lot of information, insights, and predictive value.

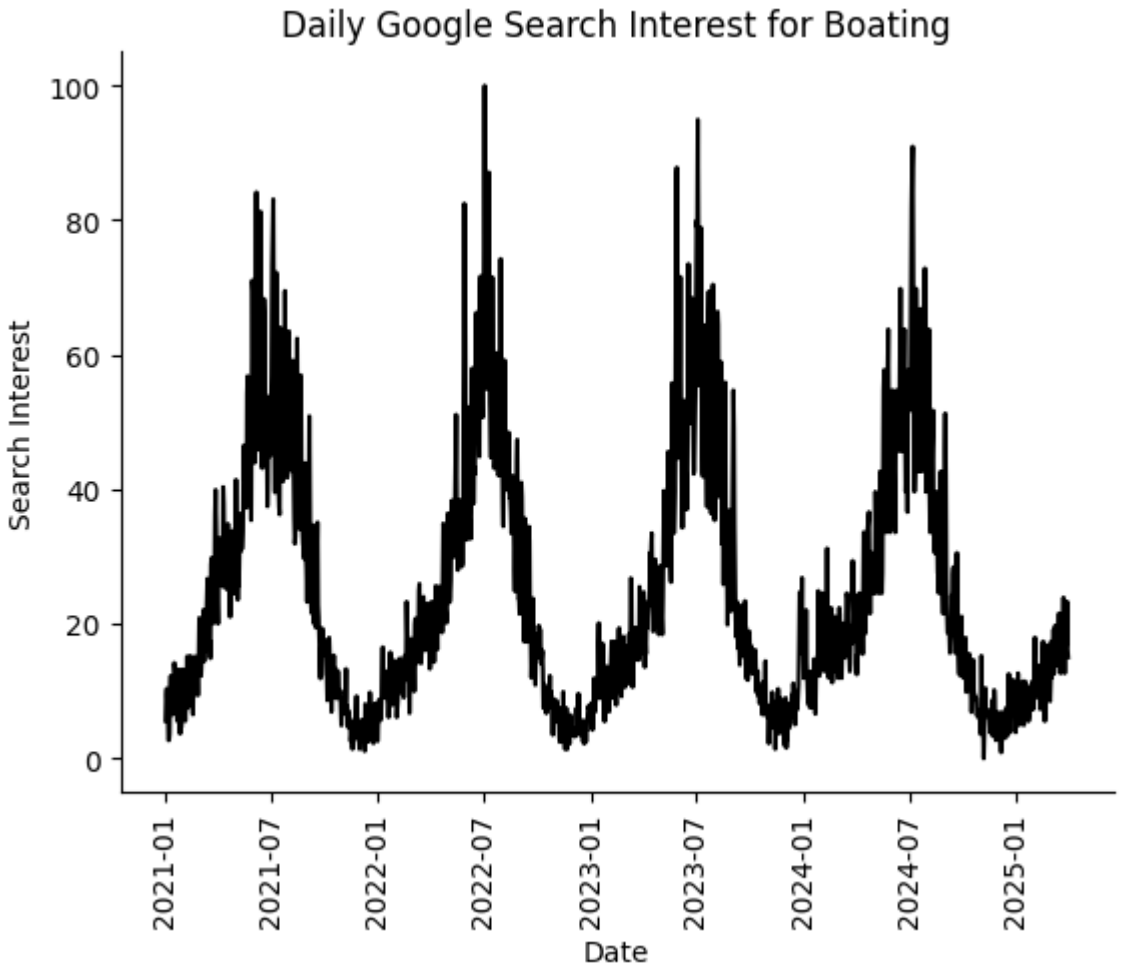
==========

Search Interest Trend Plots

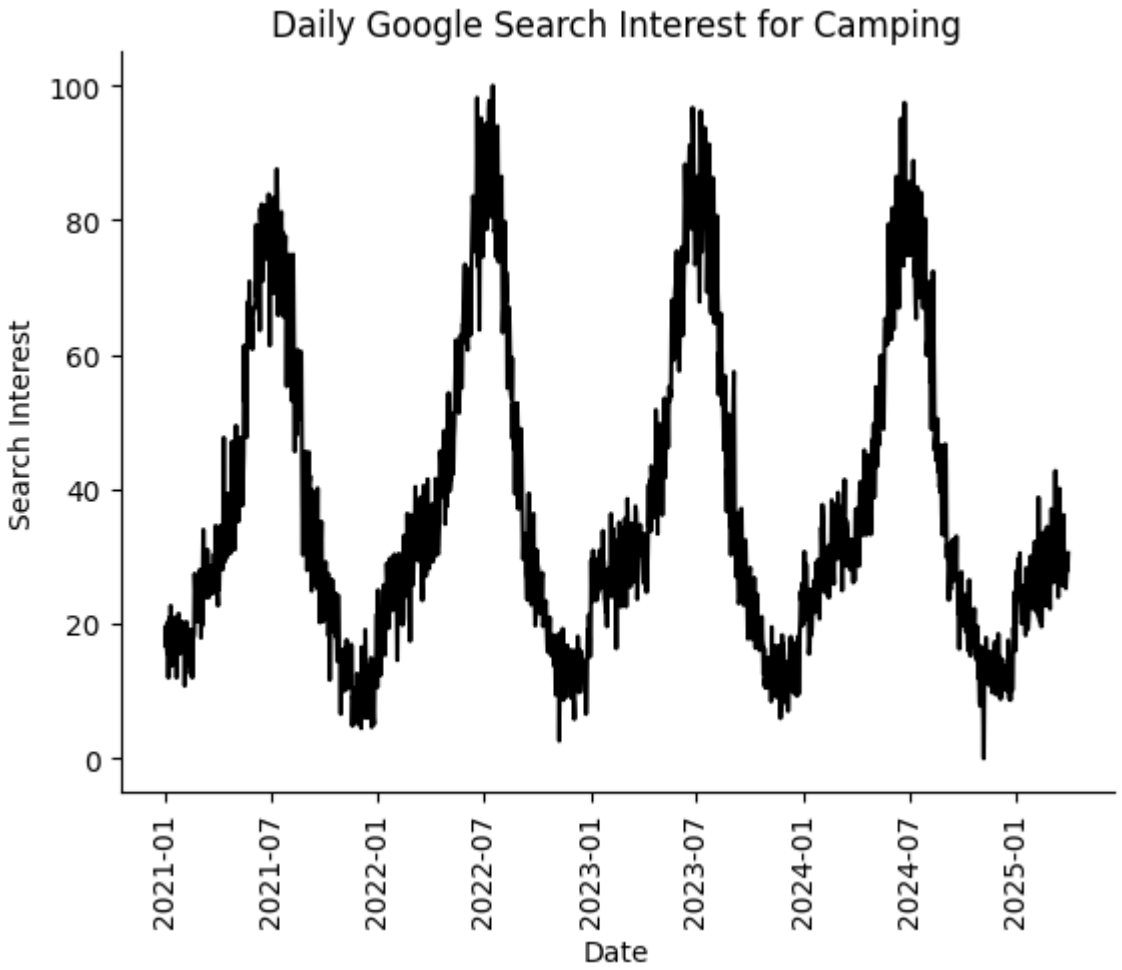
Now, I will plot the 10 resulting data series. First I plot the series for atving. As with most of these series, the plot for atving is cyclical or seasonal with most search interest in the summer months. In this case, search interest appears to be trending downward. To some extent, some forms of outdoor recreation peaked during or after the COVID-19 pandemic. This was especially the case for forms of outdoor recreation that involved buying expensive items due to the influx of government money during the same time period. ATVs were in short supply during 2020.



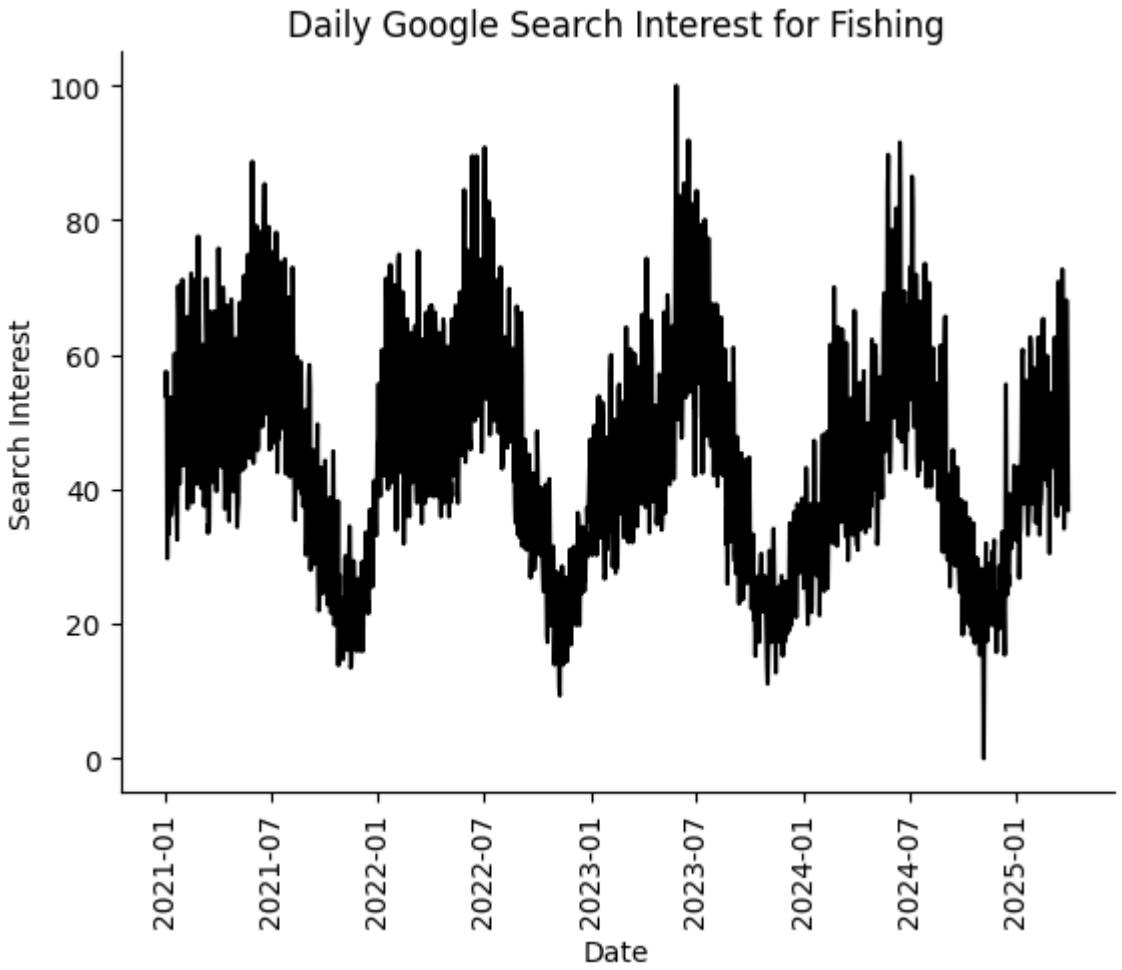

Interest for boating is even more peaked during the summer and is not trending downward unlike ATVing. Boating has very pronounced high search interest values near the 4th of July holiday.



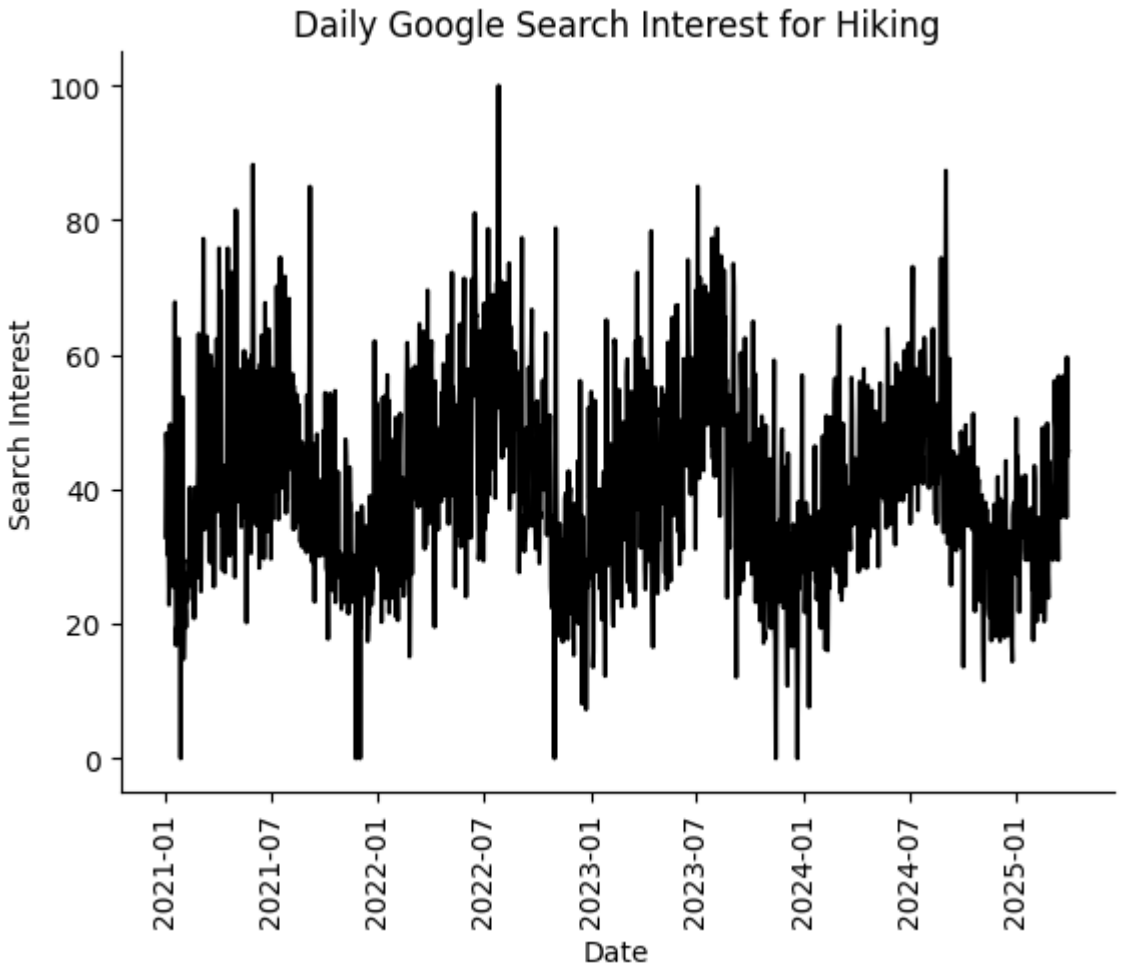

Search interest for camping has remained relatively steady overtime with much more interest in the summer than winter. Search interest is even more seasonal than boating, perhaps because camping often involves less major purchases that might occur during sales during the cold months.



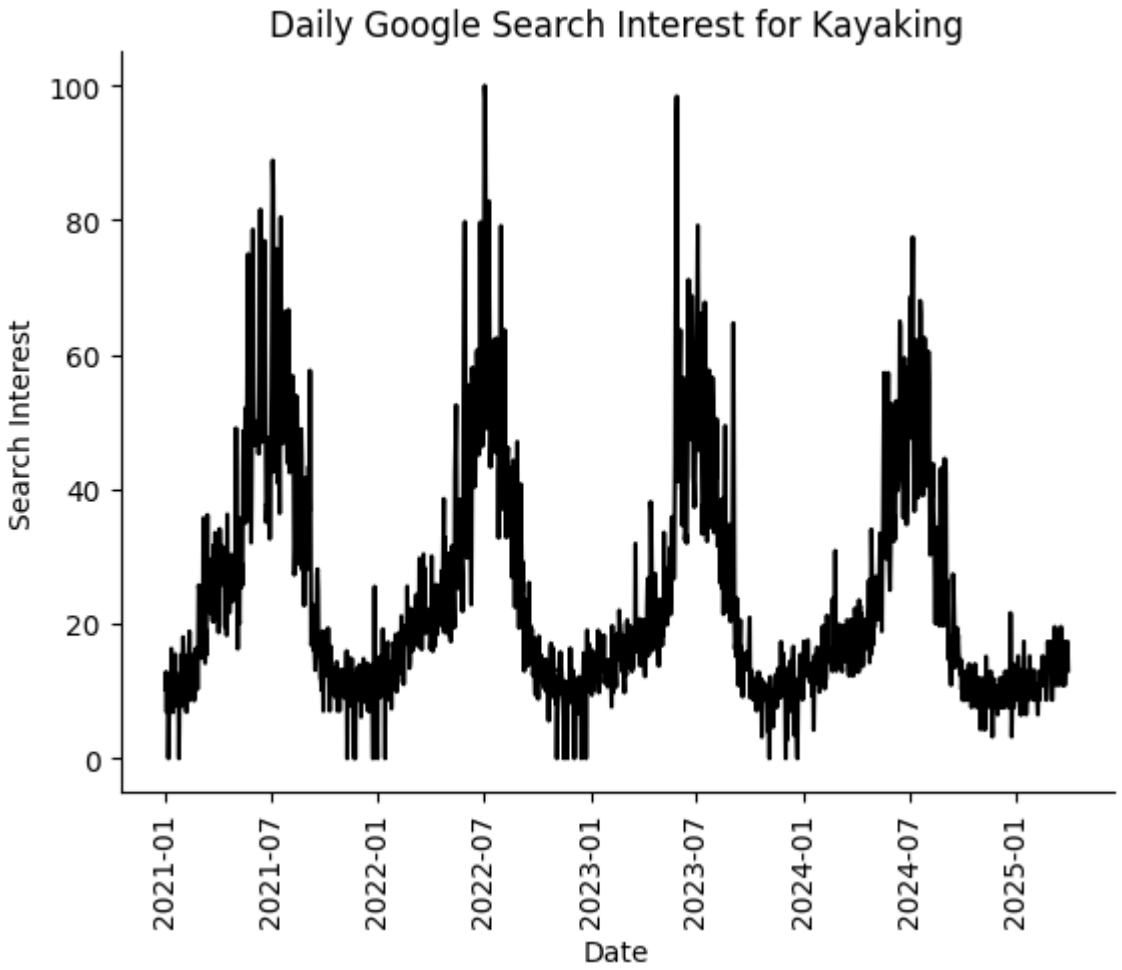

Fishing is less seasonal than other activities and has its low point in late fall. Many fishermen partcipate in ice-fishing during the winters.



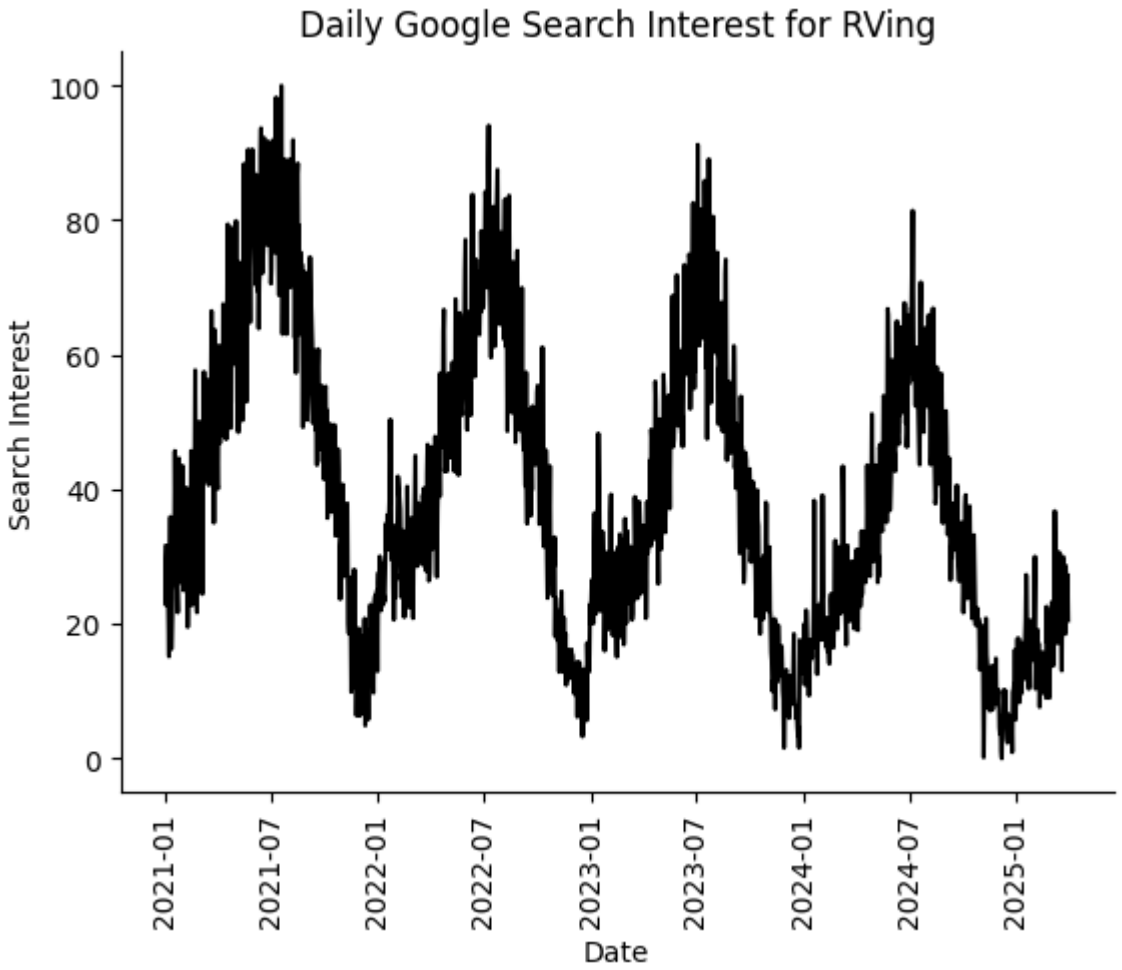

Like fishing, hiking is most popular during the summer; but hiking generates a fair amount of interest year around. Hiking is least popular in the late fall, again like fishing.



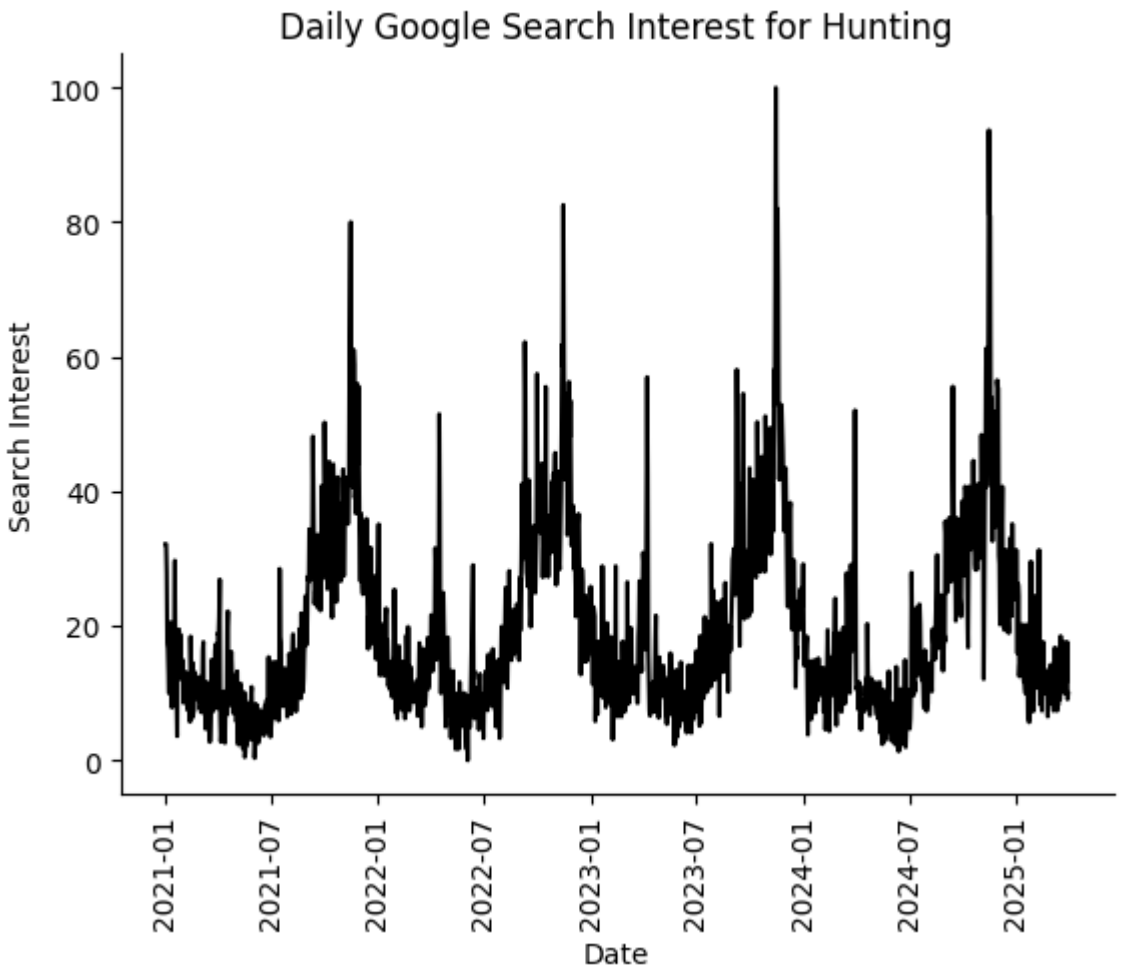

Kayaking trends are similar to boating with very high peaks and high variance during the summer. Unlike boating, kayaking does show a slight downward trend in interest.



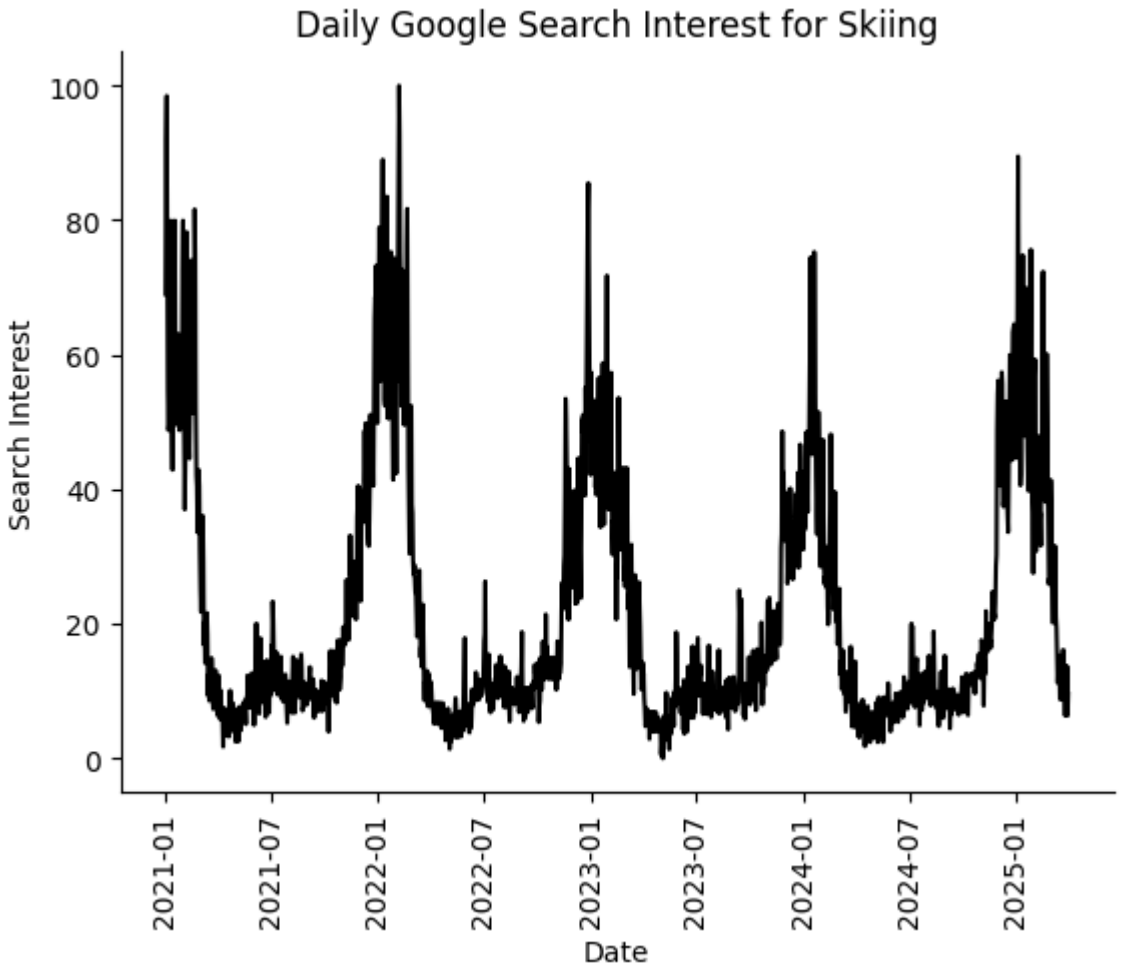

RVing is highly seasonal. While it is more seasonal than ATVing, it is simlar to ATVing in terms of declining search interest. RVs were also a popular pandemic era purchase.



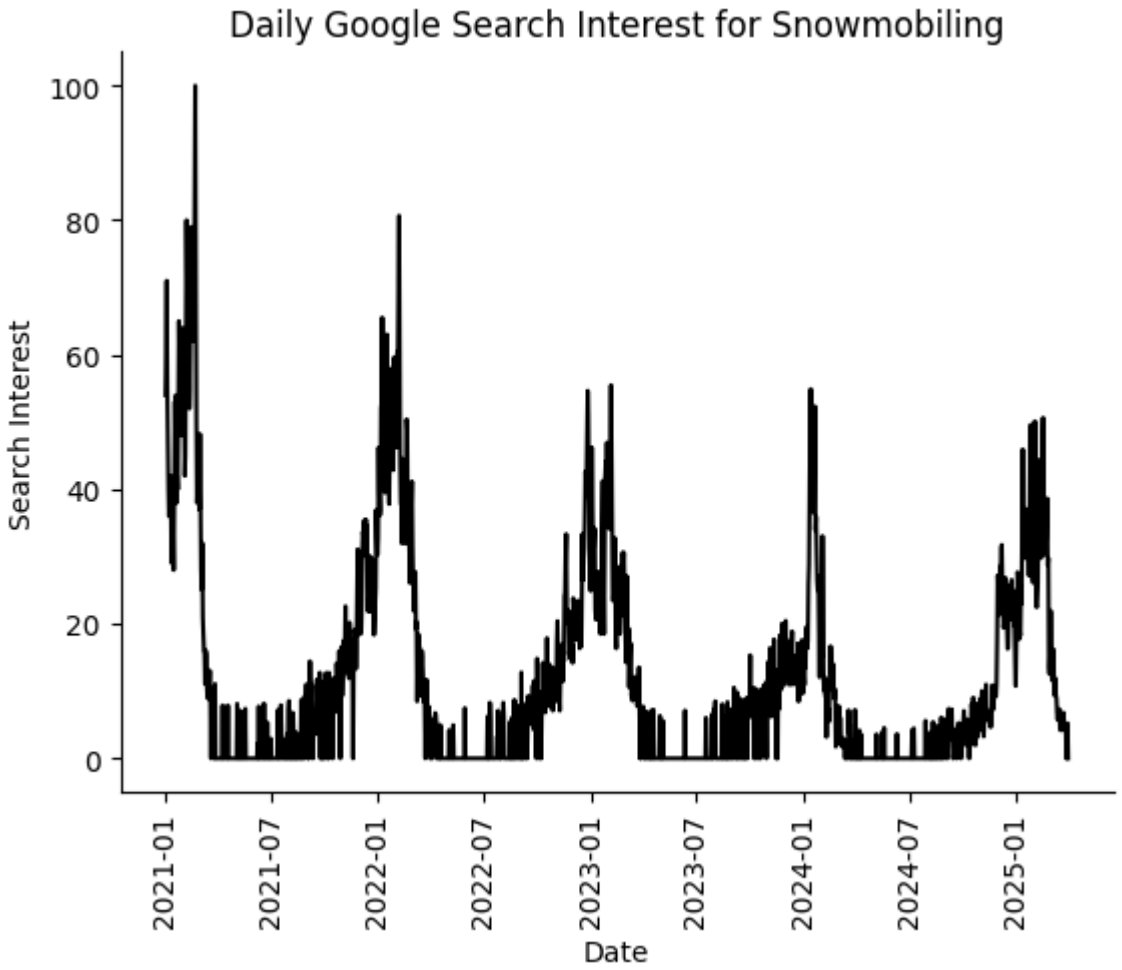

Hunting was a popular activity during the pandemic but interest appears to remain steady or even grown the last few years. Hunting search interest is seasonal in ways different from any of the other activities. Hunting is most popular during the fall, and secondarily during the spring. The very large peaks in search interest during November are on November 15th which is the first day of firearms deer season. Peaks in April are due to the start of turkey season. In modern times, these are the two most popular game species.




Unlike summer activities, skiing interest is extremely seasonal with most interest in the first few months of winter. Skiing interest was trending downward possibly due to warmer winters and less snow. However, last winter showed a return to normal winters and significant snow fall with higher search interest to match.




The final activity I analyze is snowmobiling. Snowmobiling is seasonal with interest in winter. Snowmobiling interest was trending doward, likely at least partly due to warmer winters. In fact, the 2023-2024 winter had very low retained snowfall, and search interest was extremely low. I was told that in some areas of the UP, drough condidtions were declared so businesses that rely on snow tourism could receive government support. Last winter, however, there as a return to normal snowfall and higher search interest.




==========

Search Interest Correlation Plots

Search interest between various forms of outdoor recreation is correlated. For instance, camping and boating have a correlation of over 0.9. Skiing and snowmobiling have search interest correlation of 0.9.

However, most of this correlation is spurious and due to seasonality or, in some cases, trends over time. For instance, all of the summer activties are positively correlated: atving, boating, camping, fishing, hiking, kayaking, and rving. Both of the winter activities (skiing and snowmobiling) are highly positively correlated. Hunting correlations are generally low, negative with summer activities, and slightly positive with winter activities. This is because hunting occurs in fall and spring when both summer and winter activities are less popular.

While this spurious seasonal correlation is interesting, it's not unexpected or all that useful for forecasting or prediction. The reason that hiking and kayaking appear to be related isn't due to people actually choosing these activities together but rather due to both happening to be popular during the summer due to warm weather. Changes due to seasonality can easily be anticipated by appropriate time series techniques without considering the values or correlations with other variables.

What would be more interesting and useful is to find that two activites are related even after accounting for seasonality. For instance, it's possible that many people combine activities such as camping and hiking. However, even then, correlation may be due to other factors such as responses to warmer weather or too much rain.

In the end, statistical analysis is necessary to determine more interesting results from the data. This will be the subject of the next installment.