WHY DO WE SKI? DETERMINANTS OF SKIER TURNOUT AT VAIL MOUNTAIN RESORT USING LOCATIONAL PHONE DATA

by

ASHER SAVIN

A THESIS

Presented to the Department of Economics
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

May 2022

An Abstract of the Thesis of

Asher Savin for the degree of Bachelor of Arts (or Science)
in the Department of Economics to be taken May 2022

Title: Why Do We Ski? Determinants of Skier Turnout at Vail Mountain Resort Using Locational Phone Data

Approved: *Benjamin Hansen, Ph.D.*
Primary Thesis Advisor

In this paper, I use locational phone data to model changes in skier turnout at Vail Mountain Resort and find that snow depth and cyclical factors are the primary variables driving attendance. In addition, no statistically significant effect of new snowfall on skier attendance was identified. This research both contributes to the econometric literature on the demand for ski resorts, as well as introduces the concept of using a standardized, scalable, and easily attainable data source for this purpose. As forces such as climate change and ski resort consolidation continue to affect the ski industry, there is an increasing opportunity for future research to expand this analysis nationally or globally using a similar method to the one presented here. This research could help ski resort managers and ski resort owners with a number of decision-making scenarios.

Acknowledgements

I'd like to thank my Primary Thesis Advisor, Professor Ben Hansen, for his interest and dedication to this project, along as providing the initial inspiration for it. Equally dedicated were Emily and Cameron from the data services department at the University, who helped me program, as well as talk through complex data analytics that I could never have performed on my own. Also deserving of recognition is Robert Parker, my third reader and boss at the Institute for Policy and Research Development for offering to help in any way he could and providing me with some anecdotal material for inspiration. And of course I'd like to thank the final member of my committee, Professor Lisa Wolverton, who agreed to step in at a time of need and complete the triad even though she had never met me in person.

I would also like to thank Miriam Jordan for answering my endless questions with grace, even though many of the answers were easily available online with a little effort. I'm sure she was very tired of hearing hundreds of kids ask her the same questions, but she never showed it. Trond Jacobsen, my teacher for the Prospectus course and my original CHC advisor also deserves credit for laying the foundation for my research and writing methods that I used throughout the process. Even though he was not able to remain on the committee, he helped contribute to this project in a substantial way.

Finally, I want to thank my peers, including but not limited to Aidan Potts, Katelyn McCallister, Jackson Chenoweth, and Liam Howard, for indulging me in speculation and conversation that eventually led to the ideas expressed in this paper. I could not have done any of this without them.

Table of Contents

Introduction

In 2019, ski and snowboard resorts brought in just under 3.5 billion dollars in revenue, documenting the industry's peak market size, before dropping to 2.7 billion in 2020 and 2.5 billion in 2021 due to the onset of COVID 19. However, as restrictions on the ski industry lift in 2022, industry reports forecast that ski resort revenue should jump back up to the 3-billion-dollar mark in 2022 (IBISWorld). As further evidence of a strong rebound for ski resorts, while revenue dropped in 2021, snow sport participants reached a peak of 10.5 million, 1 million of which are believed to be newcomers to their respective sports (National Ski Areas Association). Alpine ski and snowboard equipment sales also rose this season from depressed revenue numbers during the 2020/2021 season by 16 and 29 percent, respectively, showing signs of a strong push from skiers and boarders looking to get back on the mountain (NPD Group).[1] All together, these numbers indicate that while the size of the ski resort industry itself temporarily shrank during the pandemic period, more people than ever were either participating or were planning on participating in some type of winter activity, and that the ski resort industry should either return to it's pre-pandemic size or even surpass it.

Successfully managing a ski resort requires a good sense of how many skiers are likely to visit the mountain each day, week, month, or over the entire season, and if the ski resort industry does return to – or exceed – it's pre-pandemic size, it will be more important than ever for managers to do this accurately. Having a good sense of skier turnout allows resorts to hire the correct number of staff, order the correct amount of

[1] "Alpine" ski gear does not include backcountry ski and snowboard equipment, so the well-documented rise in backcountry activity during the pandemic does not influence these numbers.

food, open the correct number of trails, or make available the correct amounts of rental equipment. It also allows stakeholders to get a good idea of how much profit the resort can be expected to make over a given period of time. Historically, many managers have relied on a sense of intuition to make predictive assumptions about the number of skiers showing up each day. Over the last decade however, the ski industry has seen a shift toward conglomeration and corporatization, and this shift may have already, or will eventually result in changes to the way resorts are managed (Falk, 2009). According to data published by the National Ski Area Association, as of 2021 nine companies owned over 100 of the 462 resorts in the United States. The largest actor is Vail Resorts, which owned 37 mountains as of December 2021 (National Ski Area Association). With the increased consolidation of the skiing industry, more data-driven approaches to management will likely be sought. One of these tools will likely be – or already is – a tool for consistently predicting skier turnout.

The primary purpose of this paper is to prove the concept of using locational phone data as a viable source for future research on skier turnout. Previous econometric models predicting or identifying the determinants of skier turnout have used data collected directly from ski resorts, which can often be tedious to attain and problematic to cross-compare due to differences in collection techniques. Using locational data allows for generalizable research, as well as cheaper application of predictive modeling in the business environment. Safe Graph, a third-party purchaser of locational phone data, is the source for visitor data used in this study. The secondary purpose of this paper is more general in scope but equally as important in effect: to contribute to the growing body of literature on the ski industry by developing a new econometric model aimed at explaining the variation in daily visits at ski resorts. The model developed in this paper is focused mainly on explaining the determinants of skier turnout and less focused on predicting it, although understanding the factors contributing to skier turnout will be necessary for developing any predictive models in the future. The scope of this work is confined to Vail Mountain Resort, the most highly trafficked ski resort in America, but would ultimately stand as a case study for broader application

Literature Review

There is an existing body of research that has used econometric models to try and explain the determinants of skier turnout. Much of this research emerged from an increased interest by the academic community in the economic impacts of climate change, and therefore most models have focused on explaining the effect that weather variables have. Hamilton, Brown, and Keim (2007) were able to explain about two-thirds of the daily variation in demand at two local New England ski resorts by using snow depth, snowfall, and temperature as explanatory variables. Notably, they found that demand was not only affected by mountain conditions, but also by conditions at nearby population centers. Shih, Nichols, and Holocek (2009) found that for two Michigan resorts, snow depth had a clear positive effect on skier turnout, but that the effect of snowfall was insignificant.[2] Research conducted in Europe found that temperature thresholds have a significant effect on skier turnout (Malasevska, Haugom, & Lien, 2015), that the impact of early season snowfall was higher than later season snowfall (Falk & Vieru, 2016; Falk & Hagsten, 2016), and that elevation of resorts influences the role that snow has on demand for skiing (Falk, 2015; Falk & Vieru, 2016). Many studies have confirmed that fact that day of the week and whether or not it is a holiday have a significant effect on skier turnout as well (Malasevska, Haugom, & Lien, 2015; Hamilton, Brown, & Kiem, 2007).[3]

[2] Literature on climate change and ski resorts usually use snow depth to mean the current level of snow measured from the ground, and snowfall to mean the depth of snow that has fallen in a given period of time. For the purposes of this paper, the same definitions will be used.
[3] In 2019, Robert Steiger did an extensive review of literature on climate change risk for ski tourism, cited in the bibliography. Much of the discussion presented here is derived from his method of compartmentalizing the literature.

These studies all relied on locally collected attendance data as their source for estimating visits, although other papers have used different methods for measuring skier turnout, such as overnight stays at destination resorts (Töglhofer, Eigner, & Prettenthaler, 2011; Damm, Greuell, Landgren, & Prettenthaler, 2016), surveys of student ski clubs near local resorts (Englen & Moeltner, 2004), or winter Airbnb rental activity (Parthum & Christensen, 2022). That being said, data access has been, and still is the most limiting factor in generalizing any conclusions, since it requires direct cooperation with ski resorts or other establishments that are often hesitant or incapable of sharing data (Steiger et. al., 2019).

Safe Graph, which buys, organizes, and redistributes locational phone data from various sources, has recently received considerable attention from the academic community, especially in response to COVID-19 (Andersen, 2020; Gao et al., 2019). Safe Graph data has also been used to understand interaction patterns between racial classes (Prestby et al., 2019). One study used Safe Graph data to understand spatial and temporal visitation patterns in Florida for the primary purpose of identifying the capabilities and limitations of the data. It concluded that Safe Graph data is a viable resource for several analytical tasks and has the potential to aid industry professionals as well as city planners, especially when used in tandem with local visitor surveys (Juhasz & Hochmair, 2020). The recent yet accelerating focus on using locational phone data to measure consumer behavior indicates that this pattern will likely continue to manifest itself in future literature.

Methods

Data

Safe Graph specializes in Point of Interest (POI) data across the US and Canada. There are currently four primary dataset types: "Places", for general POI information such as location name and establishment category; "Geometry", for specific spatial boundaries and hierarchies; "Patterns", for temporal visitor patterns according to establishment; and "Spend", for transactional data. Safe Graph does not internally collect data but purchases it from a number of sources which are not publicly available.

Patterns data was used as the estimate for skier turnout in this study. Data with "Vail Mountain Resort" listed as the location name were downloaded and modified to show daily visits for the 2018/2019 and 2019/2020 season.[4] Two seasons were used since Patterns data is only available as far back as 2018, and data that might have been affected by COVID-19 was discarded. Importantly, data for the 2019/2020 season was cut off on March 8th, since COVID-19 had already started influencing public activity before the official end to the season on March 15th, as is evident from the google search trends presented in Figure 1.

[4] The 2018/2019 season spanned from November 14th to April 24th, and the 2019/2020 season spanned from November 15 to March 14. The 2019/2020 season was cut off early due to COVID-19.

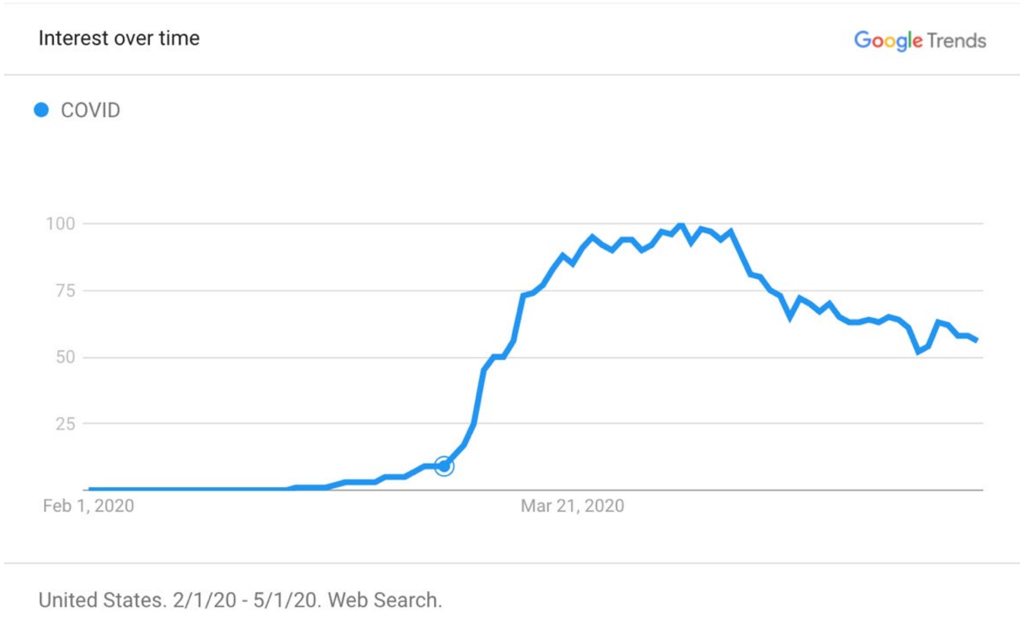

Figure 1: Google search trends for “COVID” in 2020.

The blue dot is located on March 8th, which is where the first significant leap in relevant searches first appeared.

Patterns data is collected starting at 7:00 AM for the date of interest and collection continues until 7:00 AM the following day. Patterns data is reliant on locational phone tracking, and since many people have phone privacy settings, none of the *absolute* numbers for skier turnout are accurate and results will be expressed in the form of percentages. The sampling group is assumed to be random, since there is no obvious correlation between people who would have certain privacy settings and specific types of skiers. The field for “visits” was used as the indicator for skier turnout, meaning that skiers could be double counted, but only if they exited and then re-entered the measurement area in a single day, the measurement area being the locational boundary that Safe Graph defines as Vail Mountain Resort.

All weather data was pulled from the Vail Mountain SNOTEL site. SNOTEL sites are automated snow telemetry measurement centers located in remote, high-elevation mountain watersheds in the western US. They are administered by the National Water and Climate Center, and there are over 900 stations in total. SNOTEL sites measure a variety of weather-related phenomena, including precipitation, snowpack, and temperature. Snow depth and temperature were the only SNOTEL data fields used in this study, plus a “snowfall” variable calculated by taking the daily difference in snow depth.

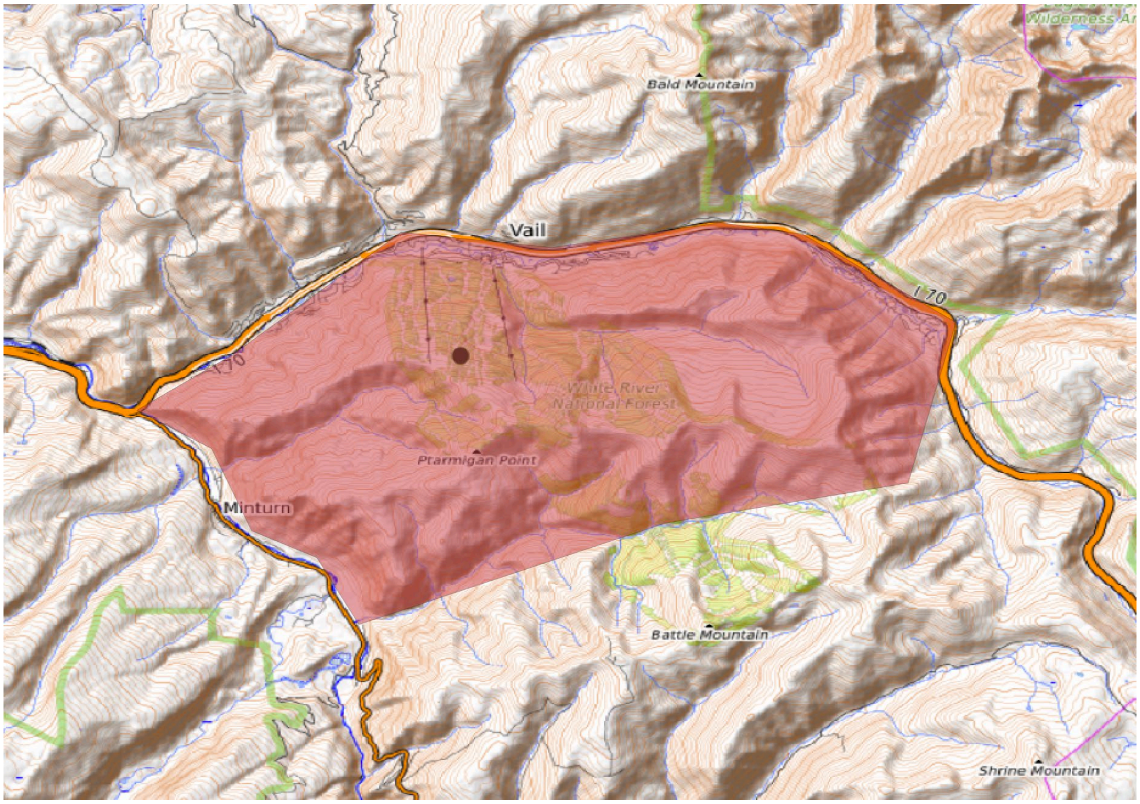

Figure 2: Safe Graph collection zone for Vail Mountain Resort and the location of the Vail SNOTEL site.

The location of the collection zone is shown in red and the SNOTEL site is represented by the black dot appearing at the top of the mountain. The polygon was constructed using the Safe Graph Geometry dataset.

Holiday data was also used in this study, collected from the Denver Public School calendars for the 2018/2019 and 2019/2020 school years.[5] Every day without class – not including weekends – was considered a holiday.

In total there were 283 observations, each representing a single day. 43 of these days were designated as holidays and 81 as weekends. Depth ranged from 7 to 74 inches, snowfall from -4 to 11 inches, and temperature from 3 to 53 degrees Fahrenheit.

[5] The Denver Public Schools calendars for the 2018/2019 and 2019/2020 school years are referenced in the Bibliography.

Model Selection

An Ordinary Least Squares regression model (OLS model) was used to investigate the determinants of skier turnout. OLS models are not normally used in time series analysis because the inherent likelihood of autocorrelation. However, just like with any data, it is appropriate as long as the data conforms to the classical assumptions required for OLS modelling. A combination of a stepwise and forward selection method was used to determine the appropriate OLS model. A stepwise selection process means all variables are explored, the highest contributing variable is added to the model, the remaining variables are explored along with the added element, the next highest contributing variable is added, and so on and so forth. Importantly, at each iteration – or "step" – the variables that were added previously are reanalyzed and removed if their significance has decreased. I decided to select my model using an altered stepwise process that permanently included or excluded certain variables when it was deemed appropriate and started with the continuous variables before adding in dummies. This part of the methodology more resembles a forward selection process.

All continuous weather variables that could plausibly have some explanatory influence over skier turnout and were possible to obtain were included in the initial stepwise iteration. These variables were: depth of snow, temperature, and new snowfall. Depth (DEPTH) and temperature (TEMP) were taken directly from the SNOTEL data, with depth represented in inches and temperature in Fahrenheit. Snowfall (FALL) was calculated using the daily difference in snow depth. Depth, temperature, and snowfall were all specified as continuous variables. Isolated regressions were run on each variable in order to determine which, if any, were significant predictors of skier turnout.

Depth and temperature were significant, with coefficients of 0.012981 and -0.012206 respectively, and p-values of 7.65e-13 and 9.08e-05. Depth had a higher R^2 term as well, of almost .16 compared to about .05 for temperature. Snowfall was insignificant with a small negative coefficient. Therefore, depth was chosen as the first variable in the model. Temperature and snowfall were then added back onto the model to see if their influence changed when paired with depth. Temperature produced similar results to its previous output, with a coefficient of -0.012206 and a p-value of 9.08e-05. Snowfall also produced similar results to its previous output, with a small negative coefficient and an insignificant p-value. Temperature was added onto the model and a regression with all three variables was executed. The output for all three remained similar to previous results, so snowfall was eliminated as a variable temporarily, leaving a regression that could be mathematically expressed as:

$$VISITS = \beta_1 DEPTH + \beta_2 TEMP$$

Control variables representing possible temporal patterns in skier turnout were then considered. Control specifications were chosen based on the patterns observed in figure 3 and figure 4.

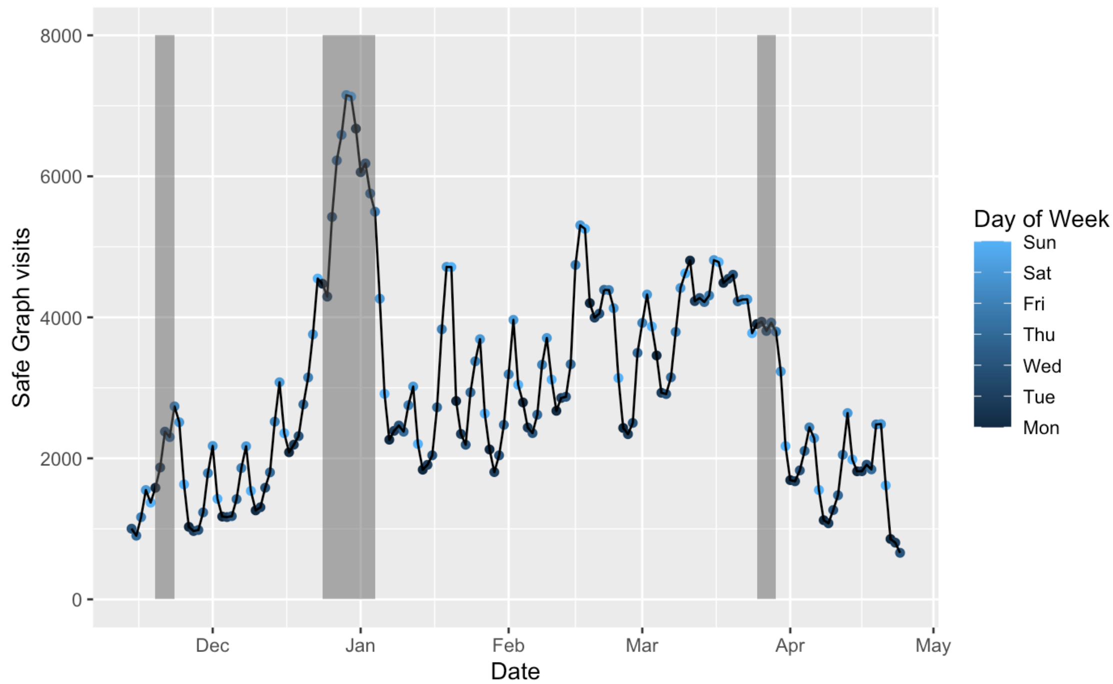

Figure 3: Safe Graph's collection numbers for "visits" during the 2018/2019 ski season at Vail Mountain Resort

Holiday Breaks for Denver Public schools are highlighted, and the observations are colored based on their position in the week.

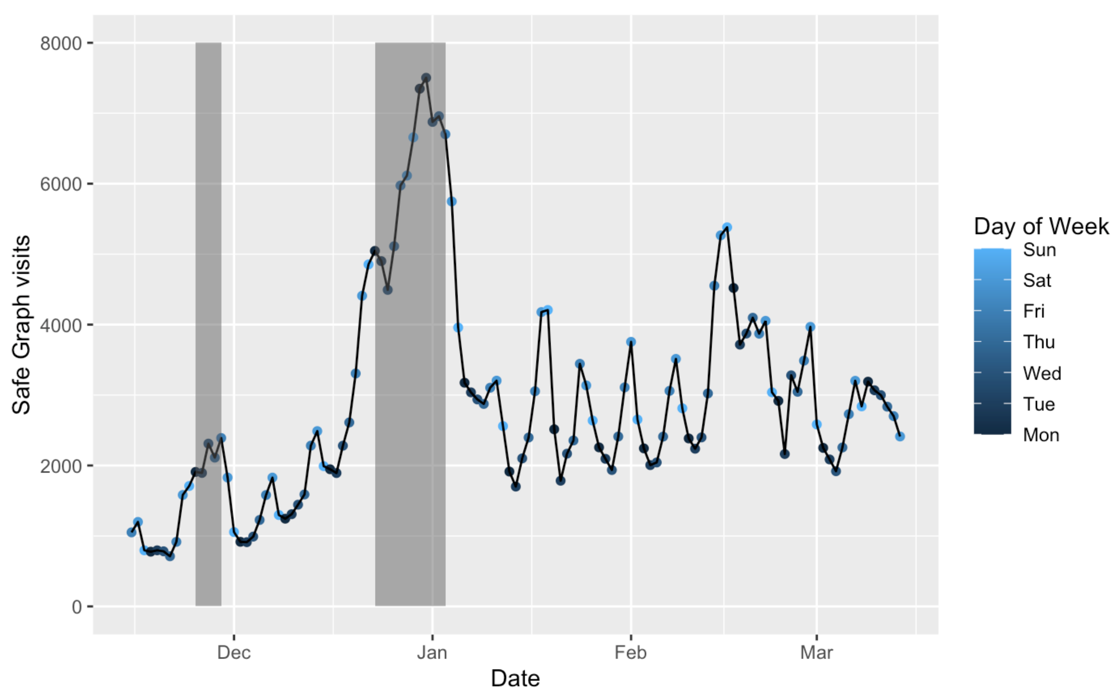

Figure 4: Safe Graph's collection numbers for "visits" during the 2019/2020 ski season at Vail Mountain Resort

Holiday Breaks for Denver Public schools are highlighted, and the observations are colored based on their position in the week.

It is clear there were consistent jumps in visits during the holidays and a consistent pattern throughout the week, with weekends – specifically Saturdays – being the busiest day. It is also clear that visits start off slow near the beginning of the season, ramp up during the middle of the season, and then decline toward the end of the season (although the decline is not evident for the 2019/2020 season because of the data specifications). There is literature that confirms that these trends replicate themselves across various ski resorts (Malasevska, Haugom, & Lien, 2015; Hamilton, Brown, & Kiem, 2007).

In order to account for these factors, three more variables were created: "month", to account for the seasonal progression of skier attendance; "day", to account for the weekly patterns; and "holiday", to account for jumps during specific holidays. Month (MONTH) and day (DAY) were specified as categorical variables according to a standard calendar, and holidays (HOL) were specified as binary variables according to the Denver Public School calendar. Each variable was added onto the existing model independently to compare their relative effect. All the added variables had significant influence over skier turnout in isolation. Day of the week had a significant effect specifically on Friday and Saturday (with Monday being the base day), and the month had a significant effect for every month except November (with April being the base month).

When the monthly indicator was added to the model, the effect of temperature was reduced to being highly insignificant (p-value of .794). This shows that the variables for temperature and month are identifying a similar trend in visits, and only one should be included in the model. Two separate models – one with month and one with temperature – were investigated. The day and holiday indicators were added to both models since both variables were found to be highly significant and did not alter any of the previous results. Therefore two models were left, expressed mathematically as:

$$VISITS = \beta_1 DEPTH + \beta_3 TEMP + \beta_6 DAY + \beta_7 HOL + \epsilon$$

and

$$VISITS = \beta_1 DEPTH + \beta_5 MONTH + \beta_6 DAY + \beta_7 HOL + \epsilon$$

When executed, these models both produced significant results for all variables included. However, the model with month instead of temperature had a much higher R^2 value of .75, compared to an R^2 value of .59. AIC and BIC tests both suggested that the model with the variable for month was preferred. This means that the time of the year itself is a better predictor of skier turnout than the temperature, even though these two variables are inherently related to one another. This leaves the proposed model:

$$VISITS = \beta_1 DEPTH + \beta_5 MONTH + \beta_6 DAY + \beta_7 HOL + \epsilon$$

To confirm that the model did not miss the effect of snowfall on visits, the snowfall variable was added back into the model one final time, and a regression with a filtered dataset excluding all holiday and weekend variables was conducted. The coefficient remained negative and insignificant, so the snowfall variable was removed from the model permanently.

Model Verification

There are six assumptions required to validate that my OLS model will be accurate and unbiased: linearity, homoskedasticity, autocorrelation, normal errors, no multi-collinearity, and exogeneity. The linearity, homoskedasticity, and autocorrelation assumptions were identified as presenting issues for the current iteration of the model, and the process of resolving each issue is addressed later in this section. The normality of errors, exogeneity, and multi-collinearity assumptions are deemed as not being disruptive to the current model, although they were addressed along with the others in order to explain why this was concluded.

The normality of errors assumption means that if the error terms were graphed, they would generally be normally distributed. This assumption usually becomes an issue when the data tends to clump around a specific observation, as in the case of an upper or lower bound. This is not relevant in this study, as there is not a single number of visitors that tends to be repeatedly expressed. A Shapiro-Wilks test was conducted on the new model in order to confirm this, and an abnormality of errors was not identified, accompanied by a p-value of .8.

Exogeneity means that there are no hidden variables which affect both our explanatory variables and our outcome variable. If this was the case, there might be another relationship which is truly driving the correlation in our model other than the one between our explanatory and outcome variables. The obvious culprit of endogeneity in this model would be the time of the season, since visitors might increase over time along with depth of snow, even if the depth of snow is not what is *driving* the number of visitors higher. This is what the dummy variable for month was included to account for. There are no other obvious perpetrators of endogeneity.

Multi-collinearity means that two independent variables are correlated with one another. This was the problem that arose from including variables for both month and temperature in the model. The stepwise process of selection was able to identify multi-collinearity in this instance, and it would have been able to identify it if there was another instance of it. The variables in the current model all have observations that exist independently of the other independent variables.

Linearity means that the dependent variable must follow a linear pattern in relation to the explanatory variables used. In this case, depth is the only non-dummy variable, so it is the only one that requires confirmation of a linear relationship with skier turnout. From figure 5, a clear linear relationship between depth and visits is obvious, especially when holiday observations are not included.

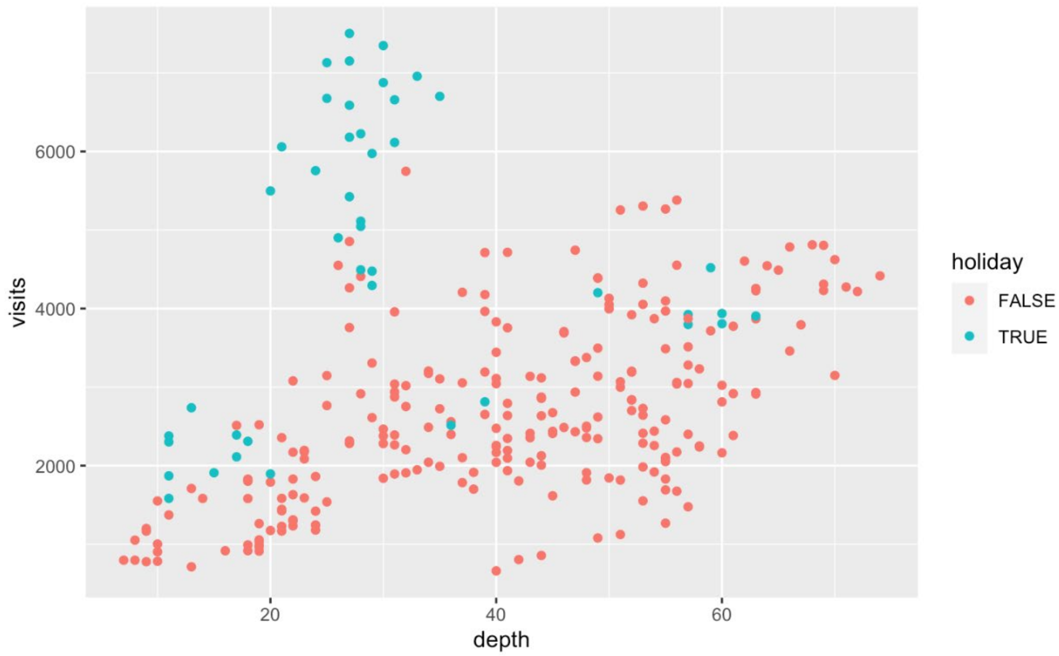

Figure 5: Relationship between depth and Safe Graph visit data

Although figure 5 shows a linear relationship, it also appears as if the error variance seems to grow larger as depth increases. This would be an issue, as it breaks the assumption of homoskedasticity, and our standard errors would not be reliable. A Goldfeld-Quant test for heteroskedasticity was conducted in order to test this and it was confirmed that heteroskedasticity did indeed exist with a p-value of about .001. In order to correct for this, both VISITS and DEPTH were logged, and the relationship is shown in figure 6.

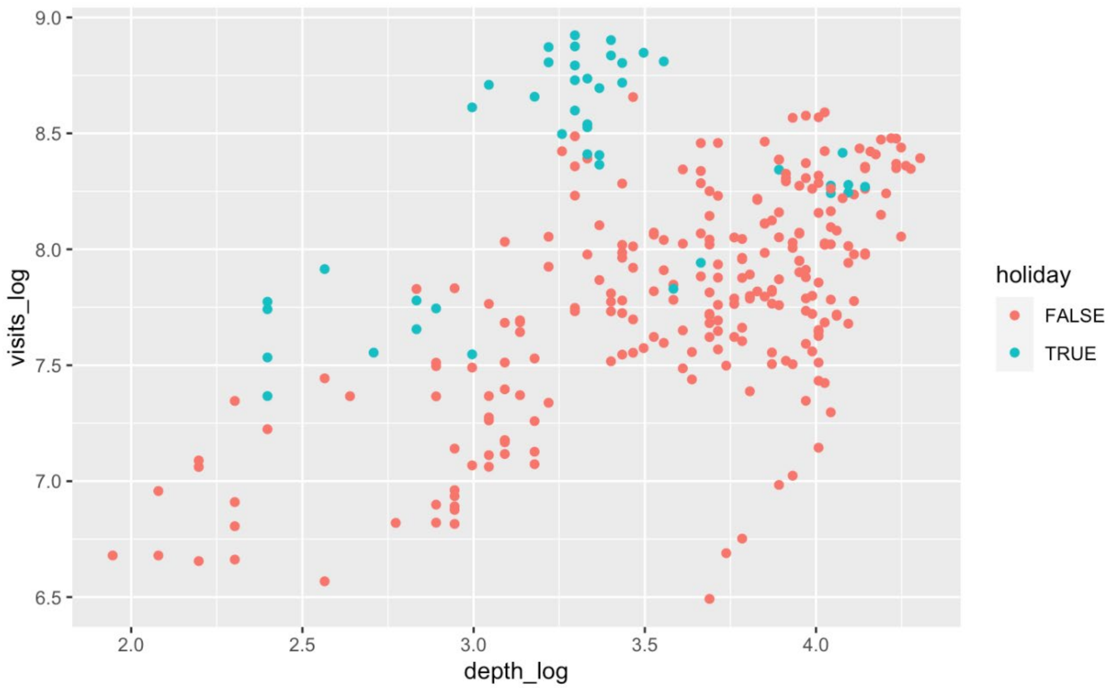

Figure 6: Relationship between log of depth and log of Safe Graph visit data

Logging both VISITS and DEPTH keeps the relationship linear while eliminating heteroskedasticity, and a Goldfeld-Quandt test confirmed this with a p-value of .49. This results in a new model, expressed mathematically as:

$$\log(VISITS) = \beta_1 \log(DEPTH) + \beta_5 MONTH + \beta_6 DAY + \beta_7 HOL + \epsilon$$

Only logging visits would have also corrected for heteroskedasticity, but the relationship would no longer have been linear.

The other assumption that complicated the validation process for the proposed model was independency of error terms, or no autocorrelation. Auto-correlation means that standard errors are related to one another, which causes significance estimates to be inaccurate. This is a particularly relevant assumption for this study since time-series are often inherently autocorrelated due to the observations being tied to one another in a temporal pattern. This was in fact the case in my model and was confirmed using a Breusch-Godfrey test. This meant is that the standard errors from the normal regression model could not be relied upon. To correct for this, a Newey-West procedure was utilized, which estimates the covariance of each term and then alters the standard errors to reflect the estimated covariance. Using Newey-West errors does not change the coefficient estimates of the regression, but increases the standard errors, and therefore decreases the statistical significance of certain estimators. Newey-West errors are incorporated into the results found in the next section.

Results

	Estimate	Std. Error	t-value	Pr(>|t|)	
(Intercept)	5.22E+00	6.11E-01	8.5419	9.95E-16	***
depth (log)	0.507598	0.148866	3.4098	0.0007499	***
December	0.541318	0.193856	2.7924	0.0056076	**
February	0.638617	0.139855	4.5663	7.56E-06	***
January	0.664645	0.155354	4.2783	2.62E-05	***
March	0.553535	0.208221	2.6584	0.0083214	**
November	0.236539	0.264392	0.8947	0.3717729	
Tuesday	0.030134	0.032923	0.9153	0.3608542	
Wednesday	0.080571	0.036047	2.2352	0.0262258	*
Thursday	0.178898	0.039956	4.4774	1.12E-05	***
Friday	0.355016	0.047263	7.5116	8.65E-13	***
Saturday	0.515903	0.048553	10.6257	< 2.2e-16	***
Sunday	0.336608	0.054993	6.121	3.28E-09	***
Holiday	0.810736	0.118798	6.8245	5.83E-11	***

Figure 7: Results from the regression model:
$\log(VISITS) = \beta_1 \log(DEPTH) + \beta_5(MONTH) + \beta_6(DAY) + \beta_7(HOL) + \epsilon$

The standard errors are corrected for autocorrelation using the Newey-West method.

Our primary variable of interest – depth (logged) – has an estimate of about .51, meaning that for every 100 percent increase in snow depth, there is a 51 percent increase in skier turnout. This estimate is significant beyond the 95 percent threshold that I will be using as my standard for significance.

January is the busiest month, with an estimate of about .66, meaning that for an average day in January, 66 percent more people are likely to show up than on an average day in April (the base month). February is the second busiest month, followed by March and then December. The opening and closing months of November and April are the least busy two months. These are all statistically significant estimates at the 95 percent level except for the November estimate, meaning we cannot conclude that November is busier than April.

Saturday is the busiest day of the week, with an expected increase in visitors of about 52 percent compared to Monday – our base day. Friday is second busiest, followed by Sunday, then Thursday, Wednesday, Tuesday, and finally Monday. The coefficient for Tuesday is the only one that is not significant at the 95 percent level, meaning we cannot conclude that Tuesday is busier than Monday. Skier turnout on holidays is expected to increase by about 81 percent compared to non-holidays. This estimate is significant at the 95 percent level.

The adjusted R squared value of the regression model was .7378, meaning that about ¾ of the daily variation in skier turnout at Vail Mountain Resort is using depth, day of the week, month, and whether or not it is a holiday. Snowfall was determined to neither be a large nor a statistically significant indicator of skier turnout and was therefore excluded from the model.

Discussion

The results of my model showed that snow depth is the primary weather variable determining skier turnout at Vail Mountain Resort, and that temperature and snowfall have either no effect, or a very minimal one. This conclusion is supported by Shih, Nichols, Holocek (2009), which concluded the same thing for two ski resorts in Michigan. Although the results from my model confirmed that snow depth is crucial in determining skier attendance, cyclical factors such as day of the week, month of the year, and whether or not it is a holiday were identified as the driving forces determining attendance. Whether or not it is a Monday vs. a Saturday has more of an effect on skier attendance (51 percent increase) than if snow depth is doubled (50 percent increase). The centrality of cyclical factors in determining skier turnout are supported by Hamilton, Brown, Keim (2007), which found that a small number of holiday and weekend days accounted for the majority of skier-days at two New England resorts. The similarity in results between my research and previous research using locally collected data helps validate Safe Graph as an effective resource for skier attendance data.

That being said, certain limitations to the data used in this paper mean that the results should be interpreted carefully. First off, Safe Graph data has the potential to capture non-skiers. Vail is not only a ski resort, but a small town with restaurants, bars, and hotels. From figure 2 it is obvious that not only the mountain itself is defined as "Vail Mountain Resort" by Safe Graph, but much of the town as well. It is likely that some proportion of people at Vail are not participating in the skiing itself, yet Safe Graph still captures them as "visitors". This effect would be especially prevalent on holidays, where people visit for an extended period of time and might take a day off from skiing. It is also possible that the sampling group is not entirely random. For instance, people's likelihood of having their locational settings set up to allow/deny locational data collection might vary by age, such as if older individuals are not as diligent about their data settings. This would cause a certain part of the population to have an inflated influence, skewing results away from representing the true population. That being said, in order for sampling error to create an issue, privacy settings must have a relationship with an outside variable, and then that outside variable must translate into different consumer choices around skiing. It is highly unlikely that this link is anything beyond spurious.

COVID-19 is also a complication when interpreting the implications of the results. As was discussed in the introduction, the pandemic caused noticeable shifts in the ski resort industry, marked by a decreased number of overall ticket sales due to resort closures at first, then a subsequent expansion of the consumer base to a broader audience. This affect is not captured in the model since the data was specifically chosen to exclude pandemic data, meaning the results might not accurately reflect the most current consumer base for Vail Mountain. The reasoning for excluding pandemic data rested on the assumption that the ski industry should return to a pre-pandemic state at some point, and therefore pre-pandemic data would be the best representation of the future state of the ski industry.

That being said, it is possible that the pandemic has permanently affected the consumer makeup of the ski industry in some way, and future research using Safe Graph data could attempt to uncover this affect. Safe Graph data allows researchers to compare pre- and post-pandemic skiers across certain dimensions, such as the distances people are driving to go skiing or the amount of time they choose to stay for. On a regional/national scale, it could also inform researchers about how skiers have shifted their resort preferences, such as whether smaller, local resorts or large, corporate ones have fared better.

Aside from the data used, the model itself had certain limitations, like the absence of certain variables that might have increased its explanatory power or focused it elsewhere. More weather variables, such as whether or not it was sunny/cloudy, the amount of time since the last rain event, or the conditions for nearby population centers – in this case Denver – might have increased the model's explanatory power. More cyclical factors, such as events going on at the mountain that might have attracted skiers, or events going on in the city that might have deterred skiers should be considered for future research as well. One of the most surprising results from the model – that snowfall did not have a statistically significant effect on attendance and that the weak correlation that was identified had a negative coefficient – could also be the result of hidden variables. Days where roads were either closed or unsafe enough to deter people were likely the days with the most snowfall, which could have been the cause for the net negative effect of snowfall on attendance. However, historical data on road conditions and closures was not accessible, and therefore this could not be accounted for in the model. The effect of snowfall could also primarily exist outside of the singular lag element – accumulation from the previous day's opening until the current day's opening – that I included in the model.

Outside of any faults in the model, snowfall may simply not have any influence in determining attendance at Vail, even though this defies common assumptions made about skiers (don't skiers want fresh powder?). The lack of a snowfall effect may be due to certain factors that are unique to Vail, such as its size, publicity, and appeal to out-of-town skiers. Many of the skiers at Vail plan their vacations months in advance and come from all over the country. For these skiers, the effect of snowfall is non-existent, since they cannot predict when the snowfall will occur; they simply decide on a time to visit based upon conveniency in their schedule. The powder-hungry locals on the other hand might prefer to ski one of the many surrounding mountains instead in order to avoid the crowds and the more expensive lift tickets. This means there may be a noticeable effect of snowfall on skier attendance at smaller, less expensive, and less publicly well-known mountains such as Copper, A-Basin, or Loveland, but not at large destination-resorts such as Vail.

The differing characteristics between various ski resorts are what made it necessary to restrict the scope of this work to a single resort. However, expanding this work to other resorts is the most interesting direction for future research, and represents the greatest advantage that using Safe Graph data has over using locally collected data. The ability to develop models derived from a single, accurate data source would allow researchers to generalize their findings by being able to rely on a standardized estimation of skier attendance. This would allow researchers to cross- compare between resorts and efficiently develop models without having to work with the resorts themselves. One application of this would be to measure which resorts' attendance is driven more by snow and weather vs. cyclical factors, which would help determine which resorts are more susceptible to climate change related threats. By extension, this would inform which resorts might benefit from investing in snow-making capabilities and which ones would not, depending on what their elasticity of skier turnout is in relation to changing weather patterns. More broadly, knowing the relative response of skiers to changing climate conditions at various resorts would help executives at large ski conglomerates make business investment decisions, such as deciding whether it would be wise to acquire a specific resort or not. There is a significant opportunity for future research to expand the methods used in this study to all the resorts Safe Graph has documentation on, thereby attaining a more generalized picture of the intricacies of the demand for ski resorts.

Conclusion

The first objective of my study was to prove the concept of using Safe Graph data to effectively model skier attendance. The alignment of my results with the results from previous, similarly focused studies using other data sources helps validate Safe Graph data as an effective resource for this use case. However, using Safe Graph data to estimate changes in attendance is not a tool that should be restricted to the ski industry. Other use cases for Safe Graph data could include analysis of consumer trends for investment purposes, internal research for companies trying to understand customer trends, or Government research related to public policy and planning. My research helps to validate Safe Graph as a reliable resource for locational data and should provide analysts in these other fields with additional evidence of that.

The other objective of my research was to explain the underlying components determining skier attendance at Vail Mountain Resort. The main conclusion from my results is that snow depth is the primary environmental factor driving skier attendance, and that together with cyclical factors such as day of the week, month, or whether or not it's a holiday, explain about ¾ of the variation in skier attendance at Vail. In a general sense, this implies that as long as Vail continues to maintain a sufficient base of snow during the ski season, skier attendance should not vary too much. Future research should attempt to uncover whether this pattern replicates itself across other ski resorts, or whether this is a pattern unique to Vail. Utilizing Safe Graph data gives researchers the tools to effectively answer this question.

Bibliography

Advanced Solutions International, Inc. “Who Owns Which Mountain Resorts.” *National Ski Areas Association*, Dec. 2022, https://www.nsaa.org/NSAA/Media/Who_Owns_Which_Mountain_Resorts.aspx.

Andersen, Martin. “Early Evidence on Social Distancing in Response to COVID-19 in the United States.”

Damm, Andrea, et al. "Impacts of +2°C Global Warming on Winter Tourism Demand in Europe." Climate Services 7 (2016). Print.

Denver Public Schools 2018-2019 School Year Calendar. http://cla.dpsk12.org/wp-content/uploads/2018/07/Calendar_1819_color-Option-Fv3.pdf.

Denver Public Schools 2019-2020 School Year Calendar. https://www.dpsk12.org/wp-content/uploads/19-20-District-Calendar-REVISED-2May2019-1.pdf

Englin, J., & Moeltner, K. (2004). The value of snowfall to skiers and boarders. *Environmental and Resource Economics, 29*(1), 123-136.

“Equipment Sales Reveal That Americans Are Back from the Backcountry and Back on the Slopes, Says NPD.” *The NPD Group*, 14 Feb. 2022, https://www.npd.com/news/press-releases/2022/equipment-sales-reveal-that-americans-are-back-from-the-backcountry-and-back-on-the-slopes-says-npd/.

Falk, M., and E. Hagsten. "Importance of Early Snowfall for Swedish Ski Resorts: Evidence Based on Monthly Data." TOURISM MANAGEMENT 53 (2016): 61-73. Print.

Falk, Martin. "A Dynamic Panel Data Analysis of Snow Depth and Winter Tourism." Tourism Management 31.6 (2010): 912-24. Print.

Falk, Martin. "Are Multi-Resort Ski Conglomerates More Efficient?" Managerial and Decision Economics 30 (2009): 529-38. Print.

Falk, Martin, and Markku Vieru. "Demand for Downhill Skiing in Subarctic Climates." Scandinavian Journal of Hospitality and Tourism 17 (2016): 1-18. Print.

Gao, Song and Rao Jinmeng and Kang Yuhao and Liang Yunlei and Kruse Jake. "Mapping County-Level Mobility Pattern Changes in the United States in Response to Covid-19." SIGSPATIAL Special 12.1 (2020): 16–26 , numpages = 11. Print.

Hamilton, Lawrence C., Cliff Brown, and Barry D. Keim. "Ski Areas, Weather and Climate: Time Series Models for New England Case Studies." International Journal of Climatology 27.15 (2007): 2113-24. Print.

Juhasz, Levente and Hochmair, Hartwig H., "Studying Spatial and Temporal Visitation Patterns of Points of Interest Using SafeGraph Data in Florida" (2020). *GIS Center*. 79.

Li, Thi. "Ski and Snowboard Resorts in the US." *IBISWorld*, Sept. 2021,

Malasevska, Iveta, Erik Haugom, and Gudbrand Lien. "Modelling and Forecasting Alpine Skier Visits." Tourism Economics 23.3 (2015): 669-79. Print.

"Natural Resources Conservation Service." *Automated Snow Monitoring*, https://www.nrcs.usda.gov/wps/portal/wcc/home/aboutUs/monitoringPrograms/automatedSnowMonitoring/#:~:text=SNOTEL%20sites%20are%20designed%20to,used%20to%20keep%20batteries%20charged.

Parthum, Bryan, and Peter Christensen. "A Market for Snow: Modeling Winter Recreation Patterns under Current and Future Climate." Journal of Environmental Economics and Management 113 (2022): 102637. Print.

"Places Data Curated for Accurate Geospatial Analytics." *SafeGraph*, https://www.safegraph.com/.

Prestby, Timothy, et al. "Understanding Neighborhood Isolation through Spatial Interaction Network Analysis Using Location Big Data." Environment and Planning A: Economy and Space 52.6 (2019): 1027-31. Print.

Shih, Charles, Sarah Nicholls, and Donald F. Holecek. "Impact of Weather on Downhill Ski Lift Ticket Sales." Journal of Travel Research 47.3 (2008): 359-72. Print.

Steiger, Robert, et al. "A Critical Review of Climate Change Risk for Ski Tourism." Current Issues in Tourism 22 (2017): 1-37. Print.

Toglhofer, C., F. Eigner, and F. Prettenthaler. "Impacts of Snow Conditions on Tourism Demand in Austrian Ski Areas." CLIMATE RESEARCH 46.1 (2011): 1-14. Print.