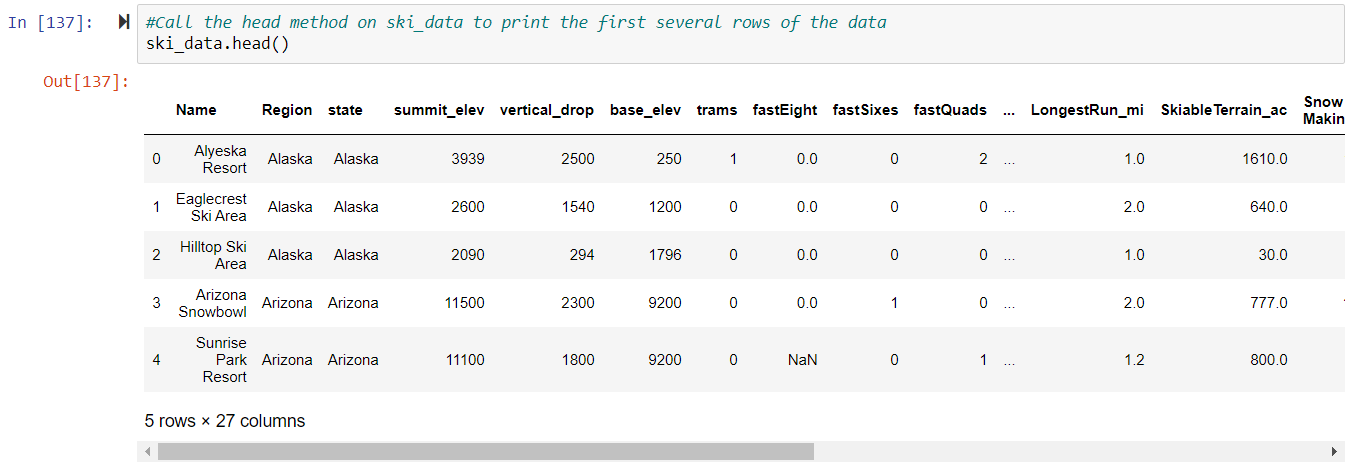
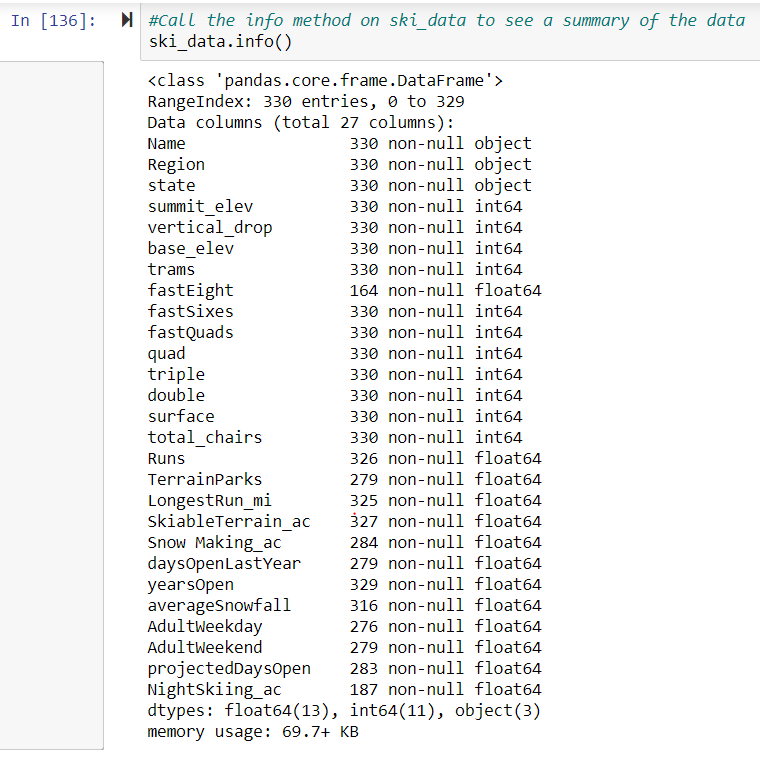
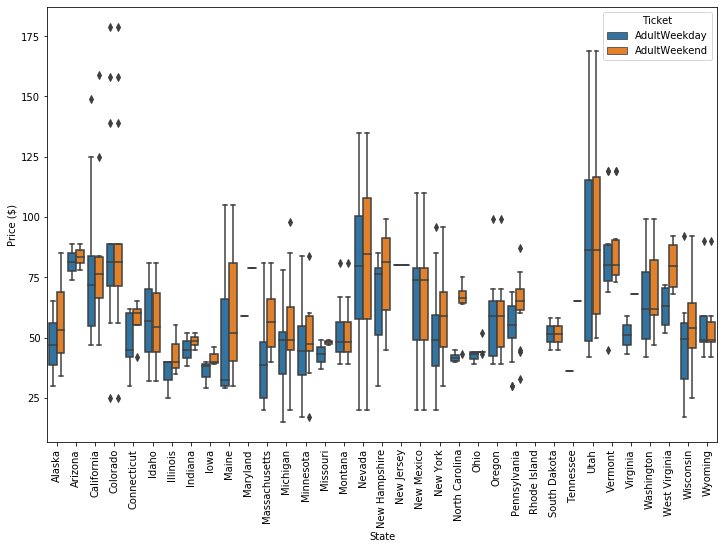
Ticket Price Modelling for Big Mountain Resort

What is the optimal ticket price for Big Mountain Resort, Montana given its unique set of features and the price-associations of features at other resorts nationwide? What features could be increased to support a higher ticket price, and could anything be removed without undermining ticket price?

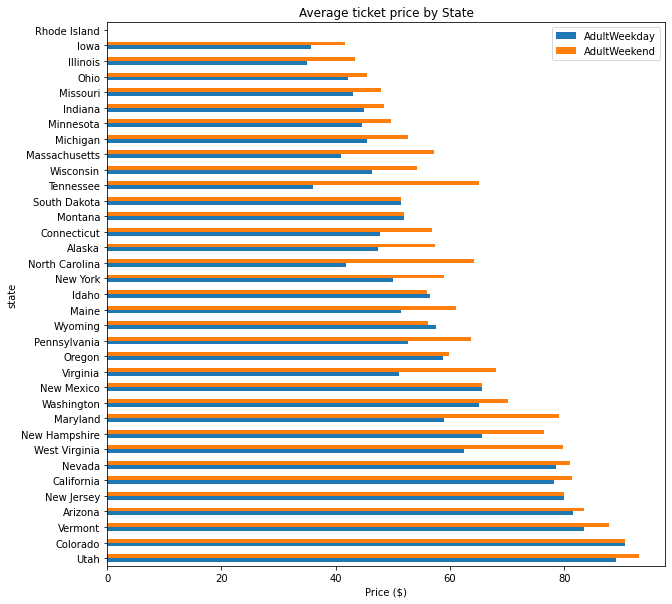




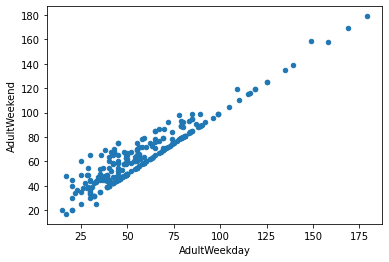
Data from the file ‘ski\_resort\_data.csv’, containing 330 rows and 27 columns. The fastEight column was dropped because half the values were missing, and the AdultWeekday column was dropped in favor of modeling the AdultWeekend column as the model’s target variable. All rows not containing AdultWeekend price values were dropped, leaving a shape of 277 rows (including Big Mountain Resort) and 25 columns.



Aside from some relatively expensive ticket prices in California, Colorado, and Utah, most prices appear to lie in broadband from around 25 to over 100 dollars. Some States show more variability than others. Montana and South Dakota, for example, both show fairly small variability as well as matching weekend and weekday ticket prices. Nevada and Utah, on the other hand, show the most range in prices. Some states, notably North Carolina and Virginia, have weekend prices far higher than weekday prices.



We've also noted another effect above: some States show a marked difference between weekday and weekend ticket prices. It may make sense to allow a model to take into account not just State but also weekend vs weekday.



We could see from this exploration to consider a few potential groupings of resorts, those with low spread, those with lower averages, and those that charge a premium for weekend tickets.

We know that two columns refer to price: 'AdultWeekend' and 'AdultWeekday'. We can calculate the number of price values missing per row. This will have to be either 0, 1, or 2, where 0 denotes no price values are missing and 2 denotes that both are missing.

missing\_price **=** ski\_data[['AdultWeekend', 'AdultWeekday']].isnull().sum(axis**=**1)

missing\_price.value\_counts()**/**len(missing\_price) **\*** 100

Out[37]:

0 82.424242

2 14.242424

1 3.333333

dtype: float64

Just over 82% of resorts have no missing ticket price, 3% are missing one value, and 14% are missing both. We will want to drop the records for which we have no price information, however, we will not do so just yet. There may still be useful information about the distributions of other features in that 14% of the data.

ski\_data **=** ski\_data[missing\_price **!=** 2]

About 14% of the rows have no price data. As the price is our target, these rows are of no use.

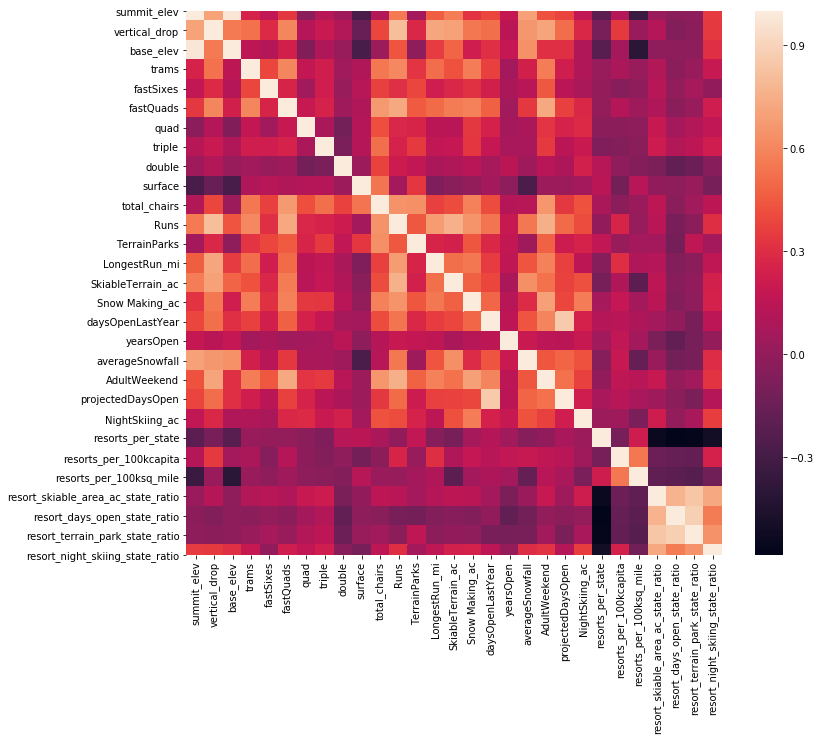
Many features in our data pertain to chairlifts, that is for getting people around each resort. These aren't relevant, nor are the features relating to altitudes. Features that you may be interested in are:

* TerrainParks
* SkiableTerrain\_ac
* daysOpenLastYear
* NightSkiing\_ac

When we think about it, these are features it makes sense to sum: the total number of terrain parks, the total skiable area, the total number of days open, and the total area available for night skiing. We might consider the total number of ski runs, but understand that the skiable area is more informative than just a number of runs.

However, we're told that we are taking all resorts to be part of the same market share, we could argue against further segment the resorts. Nevertheless, ways to consider using the State information in your modeling include:

* disregard State completely
* retain all State information
* retain State in the form of Montana vs, not Montana, as our target resort is in Montana

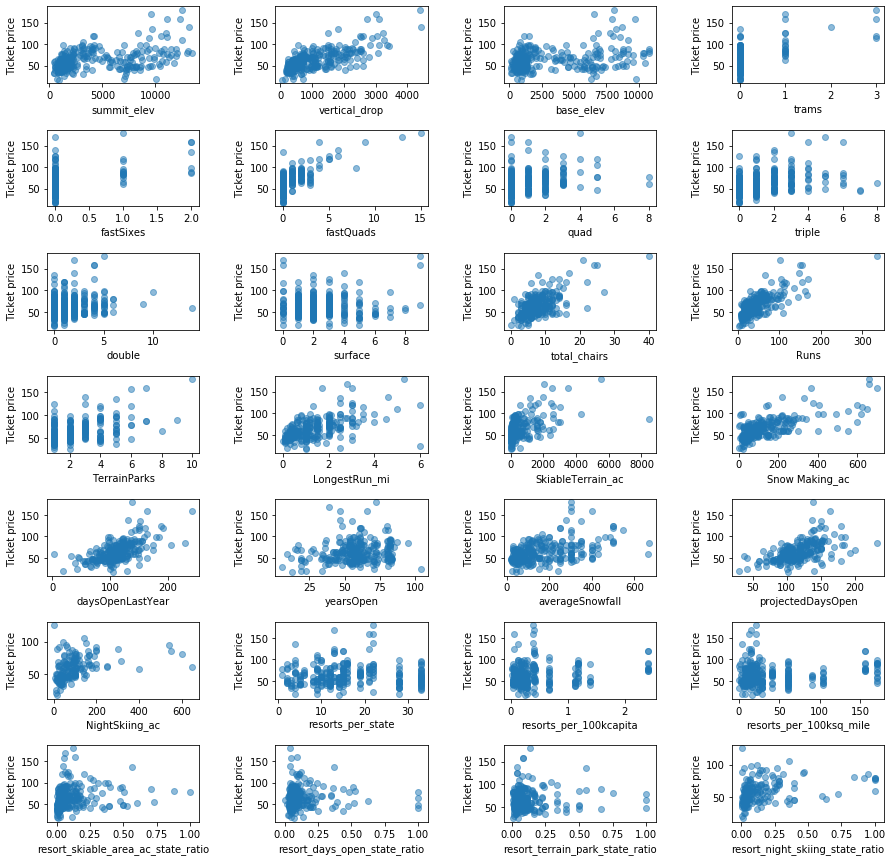


First, summit and base elevation are quite highly correlated. We've introduced a lot of multicollinearity with your new ratio features; they are negatively correlated with the number of resorts in each state. If we increase the number of resorts in a state, the share of all the other state features will drop for each. An interesting observation in this region of the heatmap is that there is some positive correlation between the ratio of night skiing area with the number of resorts per capita. In other words, it seems that when resorts are more densely located with a population, more night skiing is provided.

Our target feature, AdultWeekend ticket price, we see quite a few reasonable correlations. fastQuads stands out, along with Runs and Snow Making\_ac. The last one is interesting. Visitors would seem to value more guaranteed snow, which would cost in terms of snowmaking equipment, which would drive prices and costs up.

Of the new features, resort\_night\_skiing\_state\_ratio seems the most correlated with the ticket price. If this is true, then perhaps seizing a greater share of night skiing capacity is positive for the price a resort can charge.

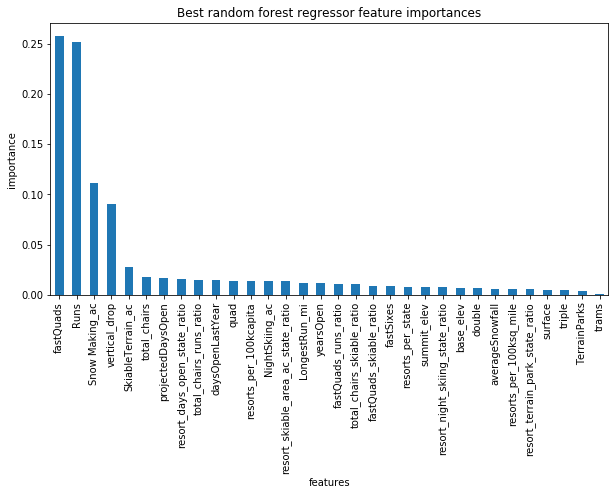
As well as Runs, total\_chairs is quite well correlated with the ticket price. This is plausible; the more runs we have, the more chairs we'd need to ferry people to them! Interestingly, they may count for more than the total skiable terrain area. For sure, the total skiable terrain area is not as useful as the area with snowmaking. People seem to put more value in guaranteed snow cover rather than more variable terrain area.



In the scatterplots, we see what some of the high correlations were picking upon. There's a strong positive correlation with vertical\_drop. fastQuads seems very useful. Runs and total\_chairs appear quite similar and useful. resorts\_per\_100kcapita shows something interesting that you don't see from just a headline correlation figure. When the value is low, there is quite a variability in the ticket price, although it's capable of going quite high. Ticket prices may drop a little before then climbing upwards as the number of resorts per capita increases. Ticket price could climb with the number of resorts serving a population because it indicates a popular area for skiing with plenty of demand. The lower ticket price when fewer resorts serve a population may similarly be because it's a less popular state for skiing. The high price for some resorts when resorts are rare (relative to the population size) may indicate areas where a small number of resorts can benefit from a monopoly effect. It's not a clear picture, although we have some interesting signs.

An auxiliary DataFrame containing state-level data was made to engineer additional features. State-level features and resort-features relative to state totals were then used to augment the original data set, leading to a total of 32 numeric features being used to predict price.

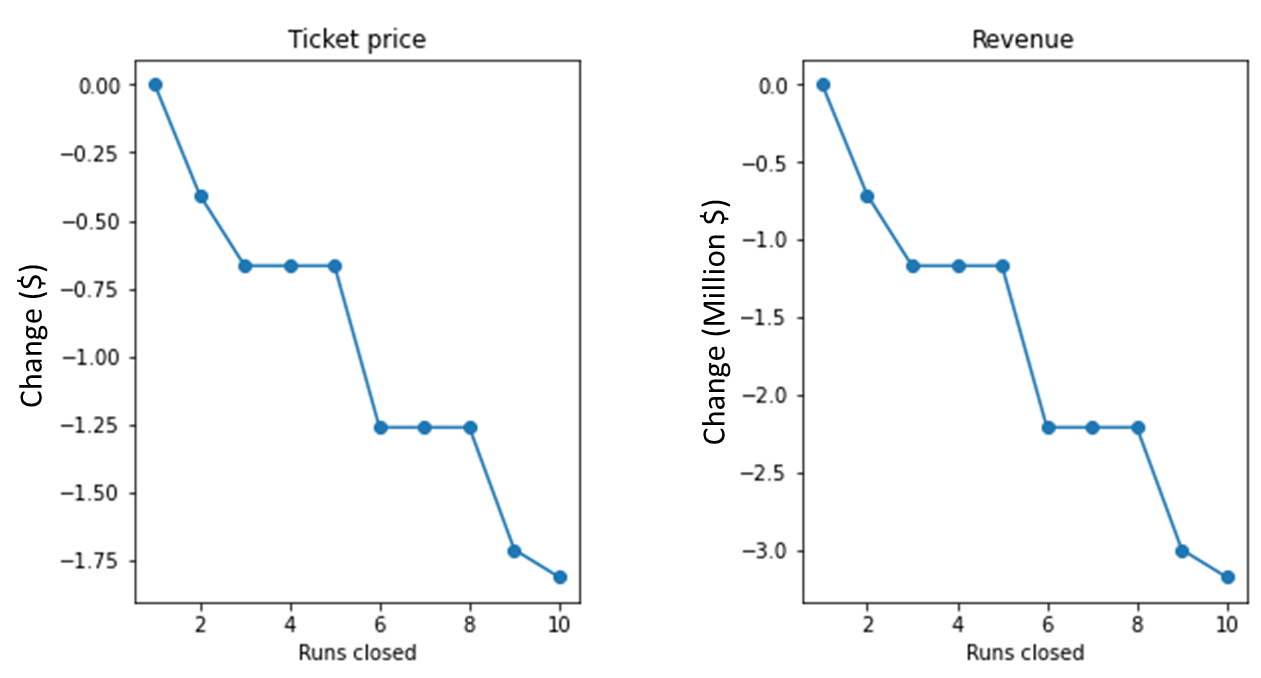
A random forest model (with 69 estimators) trained on 276 entries containing AdultWeekend ticket information revealed that Big Mountain Resort’s current ticket price is underestimating the true value.



The dominant top four features are in common with your linear model:

* fastQuads
* Runs
* Snow Making\_ac
* vertical\_drop

Considering the current features at Big Mountain Resort, **we recommend a ticket price of $95.87**, **with a range** **of $85.48 - $106.26** (+/- mean absolute error; MAE).

In addition, we tested the 4 feature alteration scenarios that were short-listed by the Big Mountain Resort team:

1. Close between 1 and 10 runs:

**Closing a single run would not change the price suggestion** but closing 2 or more would reduce the ticket price by an average of -$0.18 per run closed (shown in line plot on right).

1. Extend vertical drop by 150 feet, and add a run and chair lift:

**This would increase the price suggestion by $1.99 per ticket.**

1. Extend vertical drop by 150 feet, add a run and chair lift, and add 2 acres snow-making:

**This would increase the price suggestion by $1.99 per ticket. No added benefit over scenario 2.**

1. Increase longest run by 0.2 miles and add 4 acres snow-making.

**No gain.**

