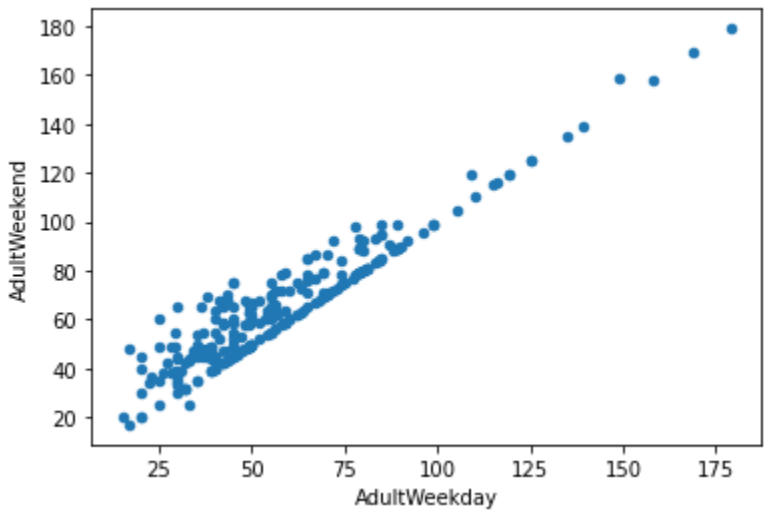
Problem statement

Big Mountain Resort can accommodate 350000 people annually. They have recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000 this season. Pricing strategy has been set to charge a premium above the average price of resorts in its market segment. It does not provide the business with a good sense of how important some facilities are compared to others and doesn't include any investment strategies options. To solve this task was proposed to analyze different ticket price strategies and look for a number of changes that could either cut costs without undermining the ticket price or will support an even higher ticket price. All analysis and modeling was based on a CSV file for 330 ski resorts from the same category operating on the US market.

Data Wrangling

Provided CSV file consisted of 330 rows and 27 columns (3 categorical (State, region and resort name) and 24 numeric columns. It required additional cleaning and checks for missing data and any discrepancies. After all required procedures, column 'fastQuad' was removed as it had more than 50% of missing data. Target value of ticket price had two corresponding columns - 'AdultWeekend' and 'AdultWeekday', cumulatively they missed about 15-16% of data. Rows with both of them missed were removed. There is a wide variability in ticket price within some states and some states have clearly very different prices to others.

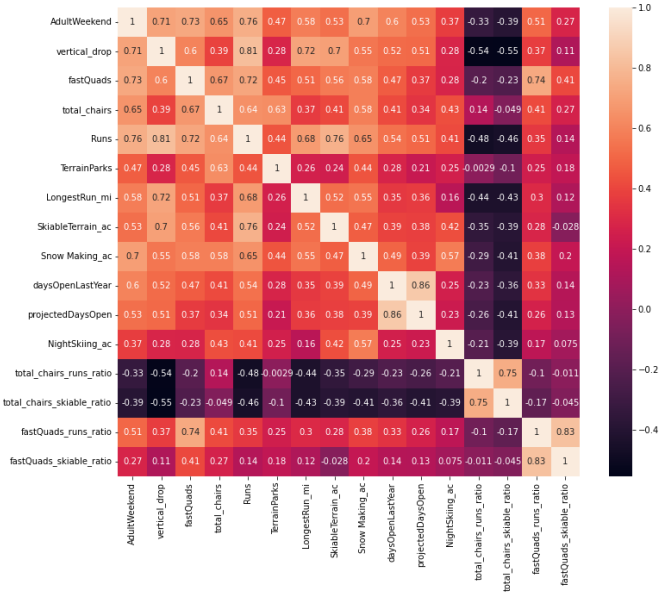


The current weekday and weekend prices for our resort, Big Mountain, were seen to be equal, at $81. This equality between weekend and weekday prices was seen for all resorts in Montana. We also saw that once the weekday price approached $100 from below, weekend and weekday prices were typically no different. Perhaps once a resort starts charging that kind of money, it's regarded as a premium resort and their pricing models don't distinguish between weekend and weekday.Thus, as we had a few more weekend prices than weekday, the latest was

removed from the table. Additional columns for the ratio of existing facilities to state wide data were added to the table.

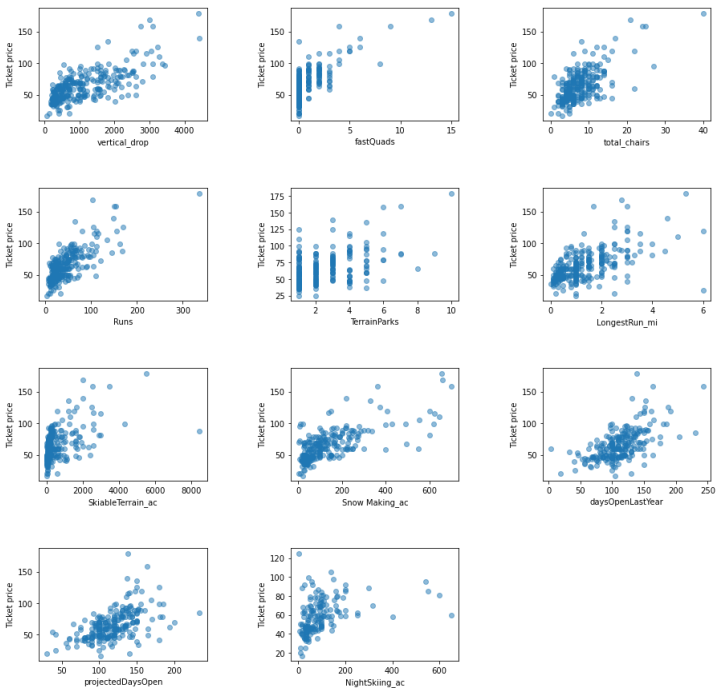
Exploratory Data Analysis

PCA analysis results gave the understanding that ticket price is not related to the State where the resorts are placed as we couldn't see any clear pattern confirming that.



There's a strong positive correlation of ticket price with 'vertical_drop' and 'fastQuads'. 'Runs' and 'total_chairs' appear quite similar and also useful for subsequent modeling. Number of snow making machines 'Snow Making_ac' can also influence ticket price as it gives perception of predictable snow coverage.

.



Ticket price modeling and possible scenarios

All data was divided to the training and testing sets in proportion of 70 to 30.

In order to make a prediction next models were built:

- DummyRegressor model with strategy='mean' was used to make a baseline model of predicting price by average. This strategy shows that mean absolute error is around $19 so you might expect to be off by this number if you set ticket price based on average of known values.
- Linear regression model was built with filling missing values by median vs mean strategies. No decisive results were found between them and probably required other transformation or subset of features to base the model on. Both train and test sets were scaled with mean=0 and unit variance. Ticket price predicted by Linear regression model could be off by $9. After running GridSearchCV looking for best linear regression model for different number of features through cross-validation it was found that best model is based on 8 features: 'vertical_drop', 'Snow Making_ac', 'total_chairs', 'fastQuads', 'Runs', 'LongestRun_mi', 'trams', 'SkiableTerrain_ac'. Vertical drop is the biggest positive feature as well as area covered by snow making and corresponds to findings during EDA.
- Additionally, a Random Forest model was built. It performed better imputing missing values with median and without data scaling, number of trees in forest - 69. Random forest model pointed to 4 important features: 'fastQuads', 'Runs', 'Snow Making_ac' and 'vertical_drop'. These are common with Linear regression model.

Both linear regression and deep forest models were compared by cross-validation mean absolute error and deep forest model performed better by almost $1. It also exhibits less variability. We could see the same consistent results on the test set.

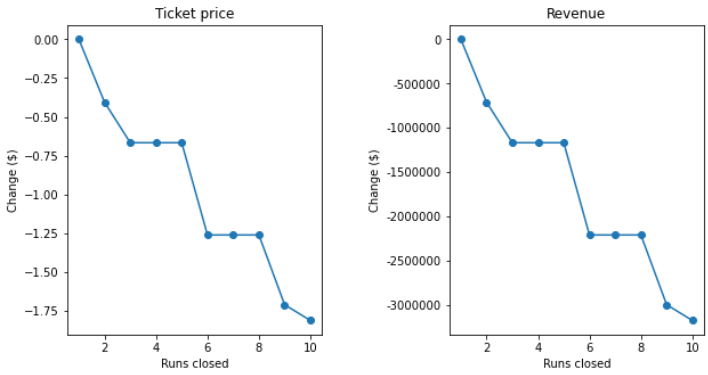
Also the quantity of data was assessed by 'learning_curve' and made sure that we have enough data for our model. There's an initial rapid improvement in model scores as one would expect, but it's essentially leveled off by around a sample size of 40-50.

Winning model was applied to the whole dataset and then to Big Mountain resort data only to obtain prediction. Predicted price is $95.87 with expected mean absolute error of $10.39, this suggests there is room for ticket price increase.

Scenarios modeling

In order to understand how facilities change can influence ticket price next scenarios were run through built model:

1. Permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics.



This scenario shows that closing one run makes no difference. Closing 2 and 3 successively reduces support for ticket price and so revenue. If Big Mountain closes down 3 runs, it seems they may as well close down 4 or 5 as there's no further loss in ticket price. Increasing the closures down to 6 or more leads to a large drop.

2. Increase the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage

This scenario increases support for ticket price by $1.99. Over the season, this could be expected to amount to $3474638

3. Same as number 2, but adding 2 acres of snow making cover

Additional snow covering didn’t give any changes to previous scenario

4. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

No difference whatsoever. Although the longest run feature was used in the linear model, the random forest model only has the longest run way down in the feature importance list.

Conclusion

Current ticket price for Big Mountain resort is $81. Model that was created before predicted price $95.87. It looks as Big Mountain resort undervalues its available facilities which stands it out as a premium segment ski resort. There is room for ticket price revision and increase even with mean absolute error of $10.39. The company is interested in researching possibilities to either cut costs without changes to ticket price or to increase ticket price in order to cover additional operational costs of $1,540,000 investments in new chair lift. Expected number of visitors this year is 350000 and the average skiing time is 5 days.

Option 1. Cutting costs by reducing number of runs:

- Big Mountain resort compares well for the number of runs as among all resorts and so for Montana state only.
- Removal of 1 run won't change ticket prices but can be efficient for cost reduction.
- Closing 2 and 3 successively reduces support for ticket price and so revenue.
- If a decision is going to be made for closure of 3 runs then the company can close 4 or 5 runs as ticket price will stay the same as with only 3 runs closure.
- Further closures could lead to an additional large drop in ticket price.

Option 2. Changes in facilities that can support ticket price increase:

- Model suggests that adding new Run, increasing vertical drop for 150 feet and adding required additional chair will increase ticket price for $1.99 that leads to increase in expected revenue by $3,474,638 (with assumptions above for number of visitors and their average skiing days)

Future scope of work

Big Mountain resort belongs to the premium market segment by facilities available. Its ticket price is competitive among Montana state resorts but lower than predicted by a model which took into consideration resorts across the US.

Further investigation can be provided if the company gives access to data related to operational costs for Runs and for Snow making facilities. Also we could see some deeper insights if we could have visitor number data across all resorts.

Visual dashboard app can be built allowing the team to see changes in ticket price and expected revenue by dynamic adjustments to available facilities and resort features.