Data Analysis for Big Mountain Ski Resort

Snow-Online.com

Bob Spoonmore

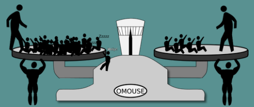
Mar 2021

Problem Statement

What opportunities exist for the ski resort, Big Mountain Resort, to increase revenue to account for the increased burden of $1.54M in added maintenance costs within the next year?

SCOPE:

Model ticket sales to find best fit for profitability



CONSTRAINTS:

- New $1.54 maint cost
- Already Predium Priced
- Current pricing based on local competitors

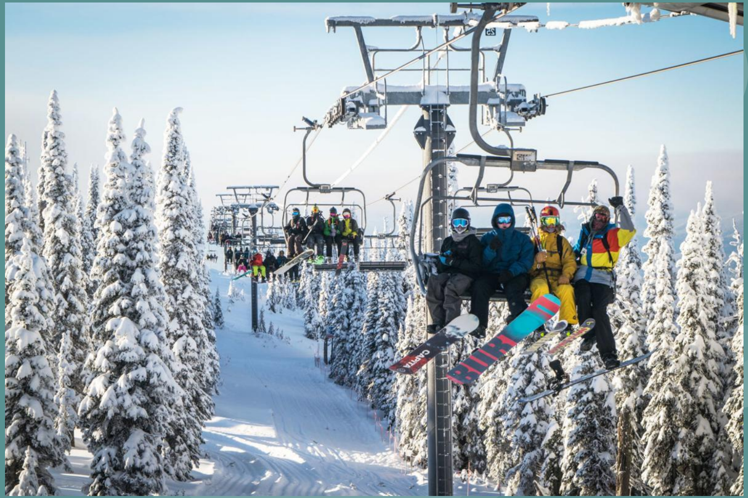
STAKEHOLDERS:

Client
Management
Leadership Team

Opportunities

Big Mountain Resort offers skiers opportunities that other resorts can't match.

- Longest run at 3.3 miles
- Skiable area at 3000 acres
- Snow Making acres at 600 acres
- Number of runs at 105
- Weekend Adult price $81 per ticket (Price is highest for Montana, but not for other resorts that compare)

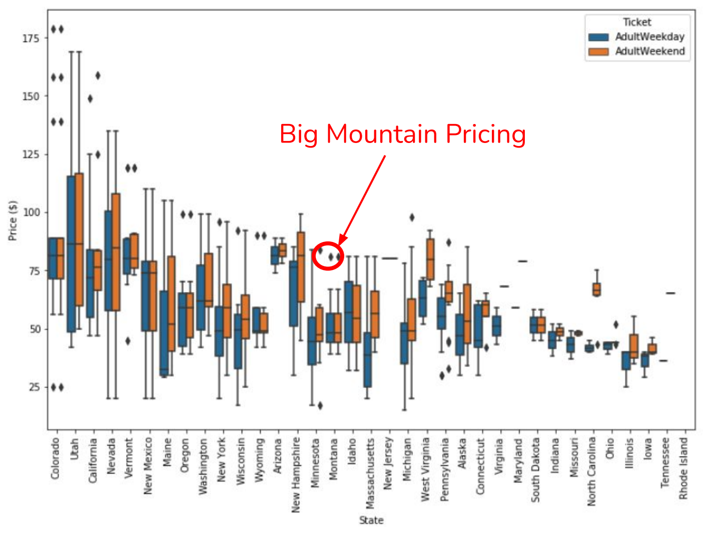


Skiwhitefish.com

How Big Mountain data compares:

- Montana ranks 10th on resort density per state
- BM 12th best for longest run distance
- BM 5th highest in Skiable acres
- BM 8th largest in Snow Making acres
- BM 20th best for number of Runs
- BM 52nd largest in Adult Weekend ticket price $$$

New York	33
Michigan	29
Colorado	22
Sierra Nevada	22
Pennsylvania	19
Wisconsin	16
New Hampshire	16
Vermont	15
Minnesota	14
Montana	12



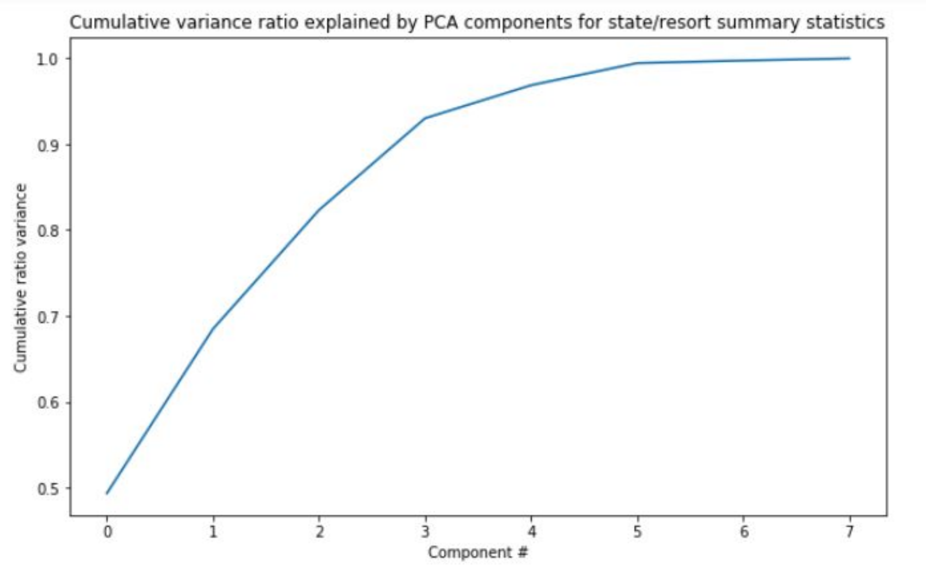
Data Wrangling:

Prepared Data
Verified no duplication
Researched extreme values in data and corrected
Removed Nan values
Removed FastEight category-too many NaNs
Joined State population data
Orig Data rows 330, cleaned to 277 unique

Exploratory Data Analysis

Principle Components Analysis (PCA)

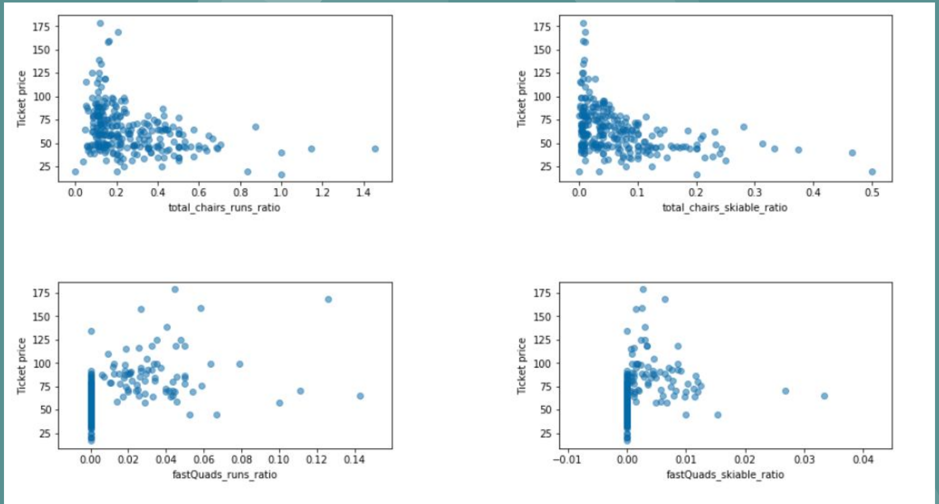
- Scale data
- Fit PCA transform
- Look at cumulative variance
    - Per graph, over 75% variance explained by 1st two components
    - 95% variance explained by 1st four components
- Look for patterns in data



Exploratory Data Analysis

Scatterplots of features

- Created scatterplot of all features
- Identified those with probably relationships
- Strong positive correlations found for features
    - Total_chairs_runs_ratio
    - Total_chairs_skiable_ratio
    - fastQuads_runs_ratio
    - fastQuads_skiable_ratio
- Other features did not initially show significant correlation to the output



Preprocessing the Data

Assessing Model Performance

- BM data removed from test data
- Train/Test split set at 70%, 30%
- Use R^2, Mean Absolute Error (MAE), and Mean Squared Error (MSE) to compare
- Linear regression applied with Dummy regressor used to fit Nan vals
    - impute missing values
    - scale the features
    - train a model
    - calculate model performance
    - After comparison, Mean and Median appeared similar

4.8.1.2.6 Assess model performance

```
r2_score(y_train, y_tr_pred), r2_score(y_test, y_te_pred)
```

```
(0.8173877345314498, 0.6704281742462251)
```

```
mean_absolute_error(y_train, y_tr_pred), mean_absolute_error(y_test, y_te_pred)
```

```
(7.623405437816813, 11.100793621683062)
```

```
mean_squared_error(y_train, y_tr_pred), mean_squared_error(y_test, y_te_pred)
```

```
(94.82179618407017, 256.97229616660246)
```

Preprocessing the Data

sklearn's Pipeline for data processing

- Applied only to training data
  - Replace missing values with median
  - Scale the data to zero mean and unit variance
  - Train linear regression model
  - Fit to training data
- Initial results appeared similar to before

4.8.2.4 Assess performance

```
In [72]: r2_score(y_train, y_tr_pred), r2_score(y_test, y_te_pred)
Out[72]: (0.8190866314532459, 0.6638181822726736)
```

And compare with your earlier (non-pipeline) result: Before using R^2 on data filled with means manually yielded: (0.8173877345314498, 0.6704281742462251) Using pipeline for R^2 yields: (0.8190866314532459, 0.6638181822726736) Values relatively close to each other.

```
In [73]: median_r2
Out[73]: (0.8190866314532459, 0.6638181822726736)

In [74]: mean_absolute_error(y_train, y_tr_pred), mean_absolute_error(y_test, y_te_p
Out[74]: (7.589995871890938, 11.16074051168558)
```

Compare with your earlier result: Before using MAE on data filled with means manually yielded: (7.623405437816813, 11.100793621683062) Using pipeline for MAE yields similar results: (7.589995871890938, 11.16074051168558)

```
In [75]: median_mae
Out[75]: (7.589995871890938, 11.16074051168558)

In [76]: mean_squared_error(y_train, y_tr_pred), mean_squared_error(y_test, y_te_pre
Out[76]: (93.9396404469242, 262.12621007050325)
```

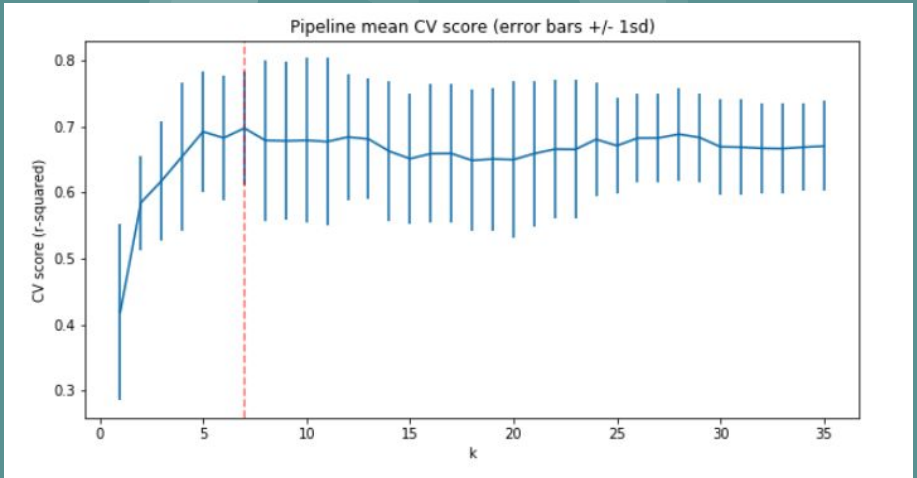
Compare with your earlier result: Before using MSE on data filled with means manually yielded: (94.82179618407017, 256.97229616660246) Using pipeline for MSE yields similar results: (93.9396404469242, 262.12621007050325)

Preprocessing the Data

Refining the model

- Used SelectBest to determine best value for K, # of subsets to slice data
- Used cross-validation to estimate model performance
- Determined k=7 best for model
- Predicted top features impacting model were as follows:

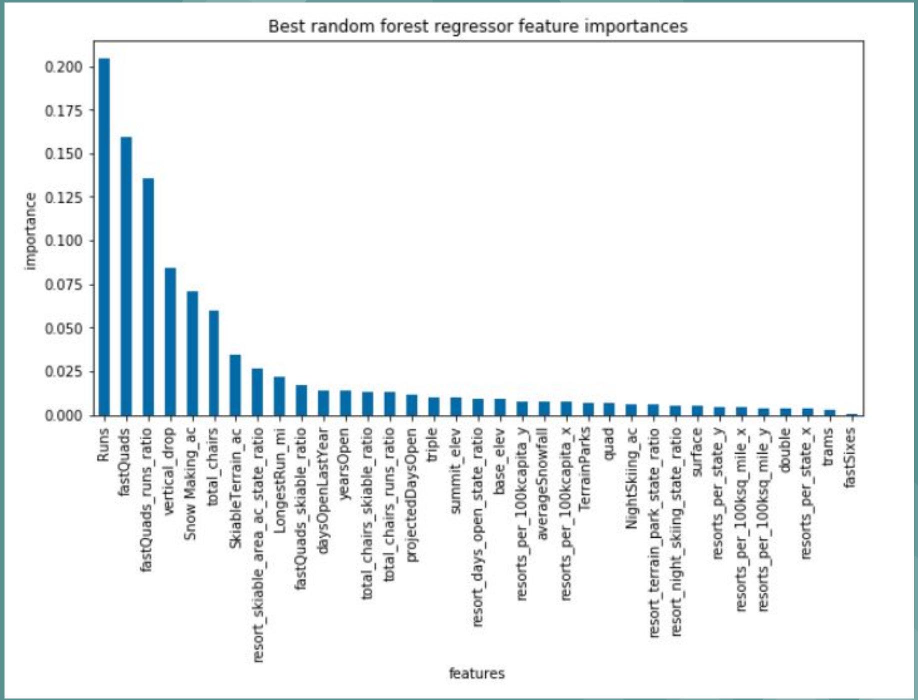
vertical_drop	8.236829
total_chairs	6.544268
fastQuads	5.769856
Snow Making_ac	4.144687
Runs	3.842936



Preprocessing the Data

Using sklearn's Random Forest Regressor (RFR)

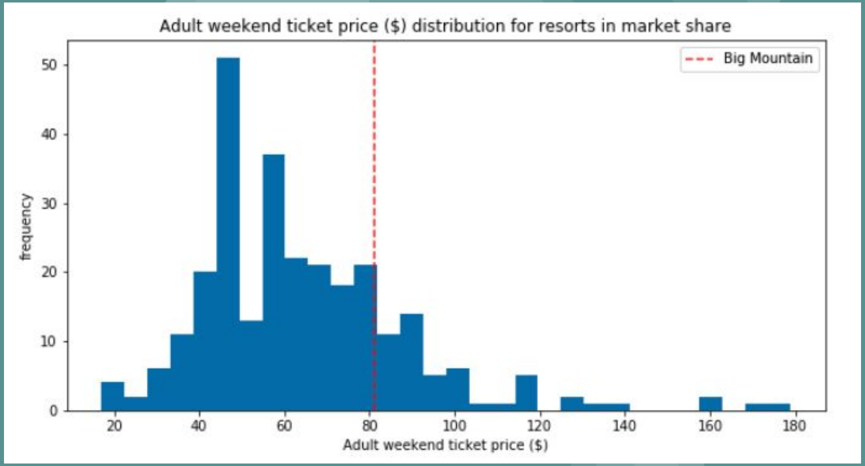
- Performs fit as part of Cross-Validation
- Tried both mean and median for hyper parameters
- Also predicted top four features to model were Runs, Fast Quads, Vertical Drop, Total Chairs, and Snow Making acres
- MSE Compare: Linear Regression 301.12 but RFR 190.2
- Choose to go forward with RFR



Modeling

Load data into model and prepare for business scenarios

- Removed Big Mountain data
- Refit data (276 rows)
- Cross-Validated model (mean 14.01, std 8.19)
- Created subplot for all important features
- Looked at the placement of our outcome value: adult ticket price. At $81 BM is above average and highest for Montana
- Reviewed the predicted top four features that impact value for ticket
- Factors determined lower impact: trams, and skiable terrain

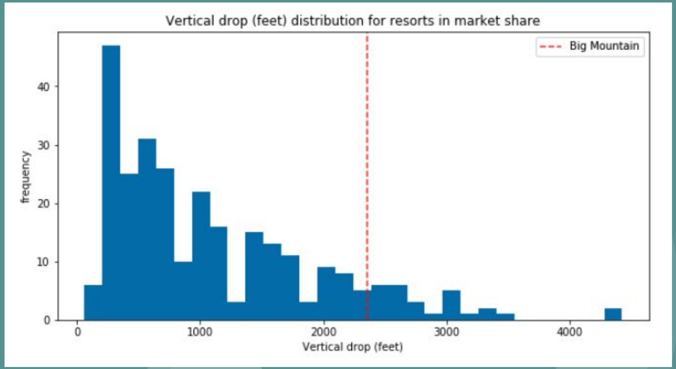


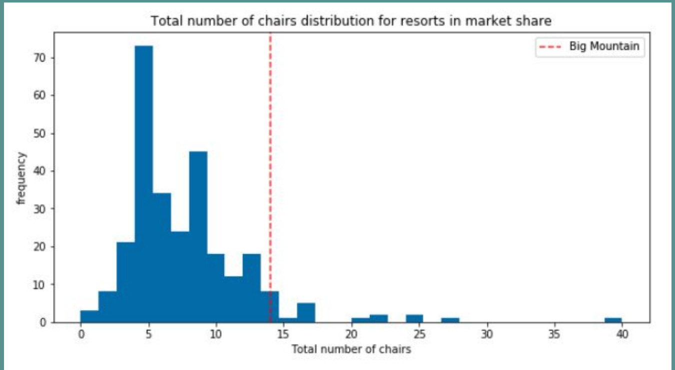
Important features to consider

- vertical_drop
- Snow Making_ac
- total_chairs
- fastQuads
- Runs
- LongestRun_mi
- trams
- SkiableTerrain_ac

Modeling

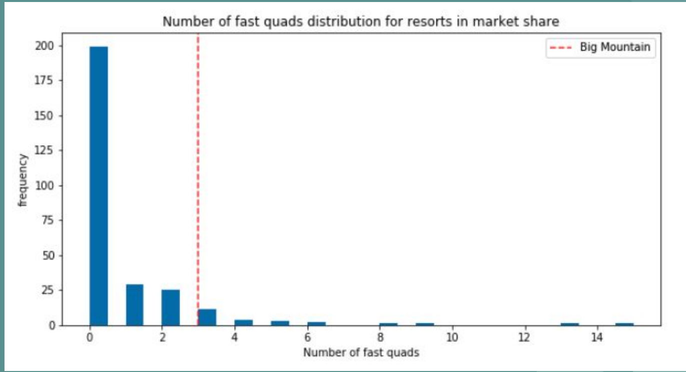
Top predicted feature 1: Vertical Drop
Above average, but not leading





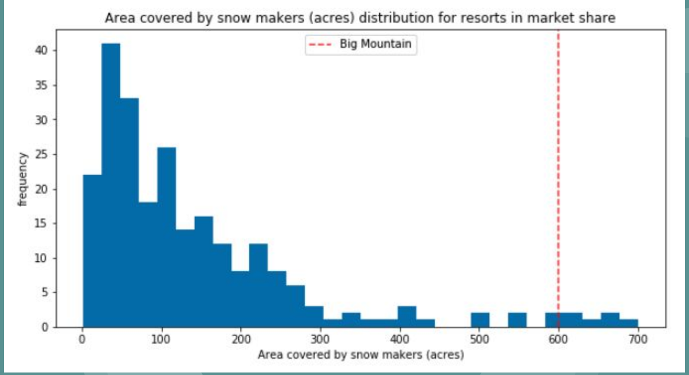
Top predicted feature 2: Total Chairs
Significantly above average, few better

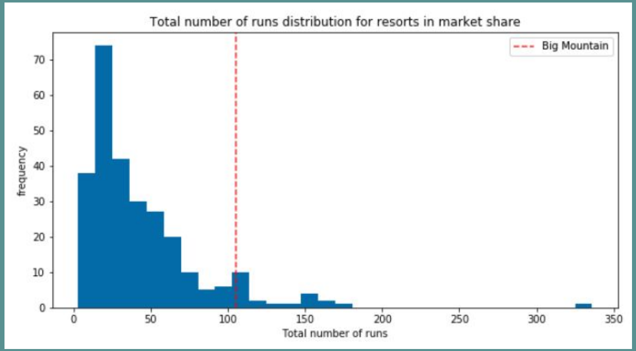
Top predicted feature 3: Fast Quads
Close to the best



Modeling

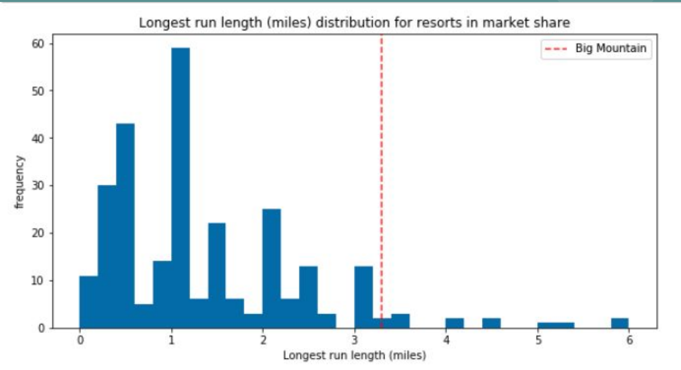
Top predicted feature 4: Snow Making Acres
Significantly above average, one of the best





Top predicted feature 5: Number of Runs
Significantly above average, few better

Other predicted feature: Longest Run
Above average, Close to the best



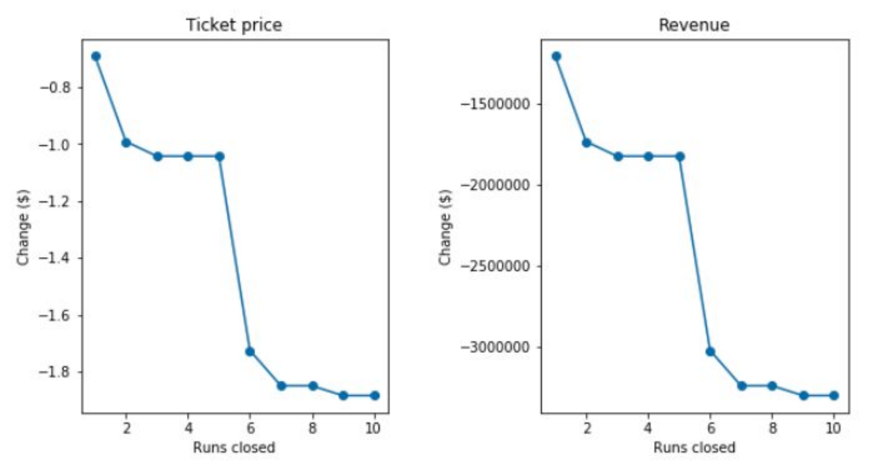
Business Scenarios

S1: Close up to 10 of the least used runs

Model Prediction:

All results predict a revenue LOSS

- Close 1 little impact
- Close 2,3,4,5 about same will bear a ticket price $1 lower but reduce revenue close to $2M
- Close 6 or more runs, significant ticket price impact of about $2 with a $3M revenue loss



Business Scenarios

S2: Add a run, increase a vertical drop by 150 feet, add a chair lift

Model Prediction:

- Model inputs

```
ticket2_increase = predict_increase(['Runs', 'vertical_drop', 'total_chairs'], [1, 150, 1])
revenue2_increase = 5 * expected_visitors * ticket2_increase
```

- Model Outputs

```
This scenario increases support for ticket price by $1.17
Over the season, this could be expected to amount to $2042918
```

Business Scenarios

S3: Add a run, increase a vertical drop by 150 feet, add a chair lift, and add 2 acres of snow making

Model Prediction:

- Model inputs

```
ticket3_increase = predict_increase(['Runs', 'vertical_drop', 'total_chairs', 'Snow Making_ac']
revenue3_increase = 5 * expected_visitors * ticket3_increase
```

- Model Outputs
  Results same as S2

```
This scenario increases support for ticket price by $1.17
Over the season, this could be expected to amount to $2042918
```

Business Scenarios

S4: Increasing the longest run by .2 miles, guarantee snow coverage by adding 4 acres of snow making capability.

Model Prediction:

- Model inputs

```
predict_increase(['LongestRun_mi', 'Snow Making_ac'], [0.2, 4])
```

- Model Outputs
  Results show ticket price change $0
  no impact on ticket prices, no impact on revenue

Conclusion

Given the problem statement: What opportunities exist for the ski resort, Big Mountain Resort, to increase revenue to account for the increased burden of $1.54M in added maintenance costs within the next year?

A Random Forest Regressor model was identified to have the best accuracy. Fitting the data for resorts comparable to Big Mountain, and merging population and size information provided the best fit model. All business cases applied to model

4 Business cases considered: **Case #2** (Add a run, increase a vertical drop by 150 feet, add a chair lift) provided the best projection for ticket price adjustment. The increase of $1 to ticket price projects revenue gain by $2M which achieves the Problem Statement. Case #3 produced same ticket price justification, but required more capital costs

https://www.spokesman.com/blogs/outdoors/2017/apr/11/whitefish-mountain-confirms-banner-year-ski-resorts/