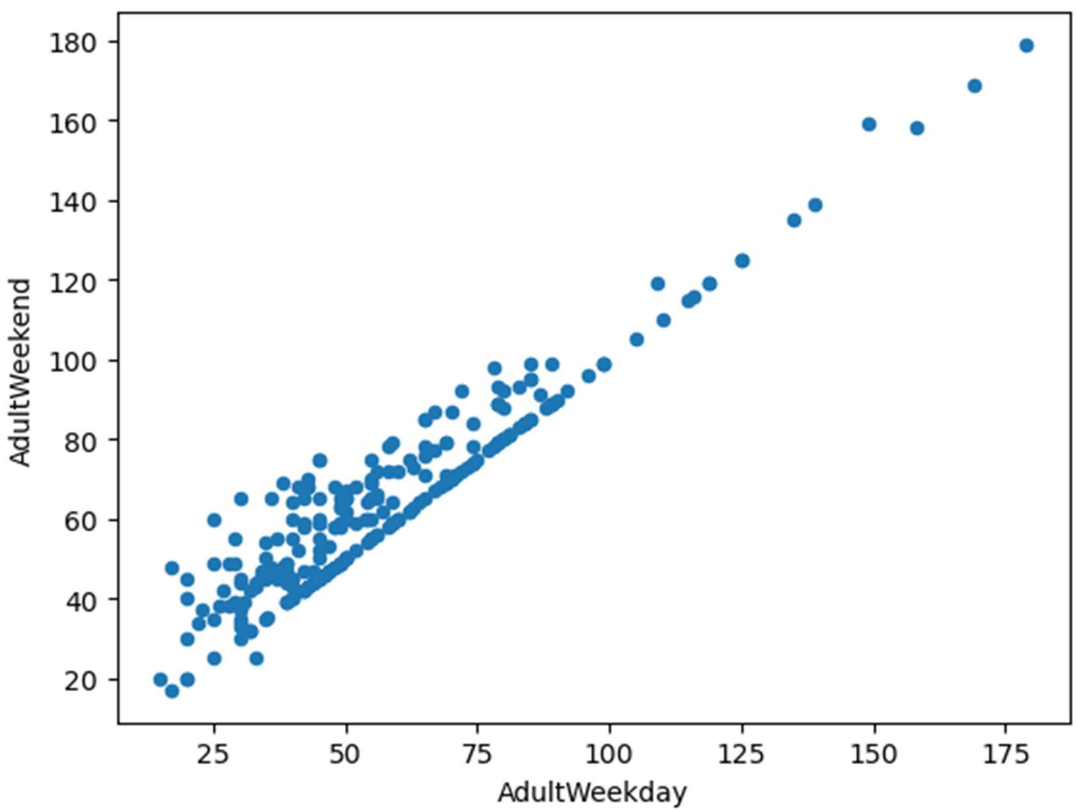
Problem statement

Big Mountain Resort is a ski resort located in Montana. The resort's pricing strategy has been to charge a premium above the average ticket prices of resorts in its market segment. This strategy is not adequate to efficiently capitalize on its facilities. The resort wants to implement a more-data driven business strategy that selects better value for their ticket price and leads to changes by cutting costs without undermining ticket prices or by charging a higher ticket price. What is the optimal ticket price that Big Mountain Resort should charge its customers to increase its profits by at least 5% in the coming year?

Data wrangling

The original data contained 330 rows and 27 columns. We have removed the 'fastEight' column because over 50% of its values are missing and the remaining values are almost all zero. We have also removed the 'AdultWeekday' column because although its values are closely related to the 'AdultWeekend' column (as shown in the plot below), it had more missing values. We chose 'AdultWeekend' as our target feature.

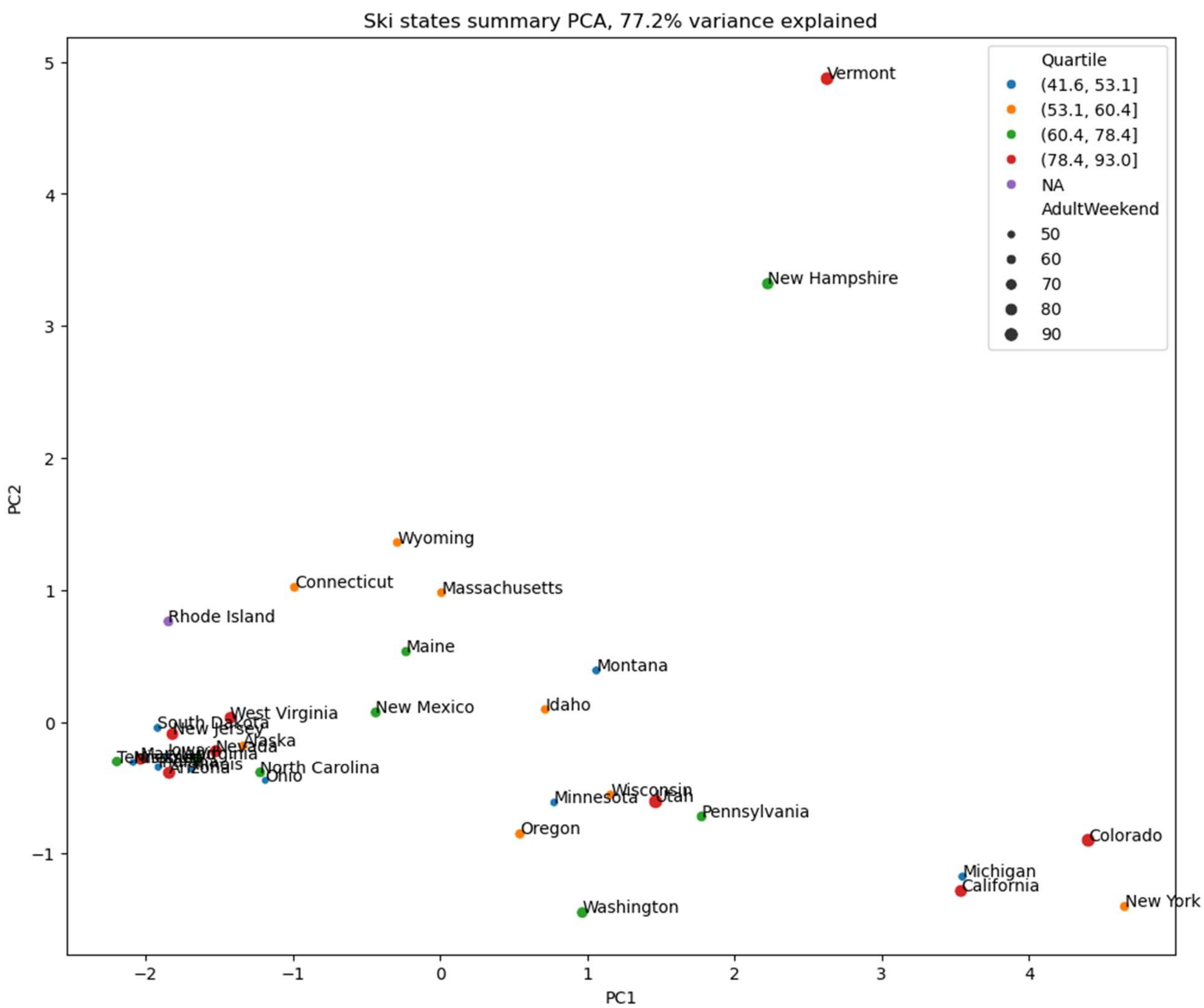


We have also removed several rows for which we have no ticket price information. We encountered some outliers and acted by correcting data for the large 'SkiableTerrain_ac'

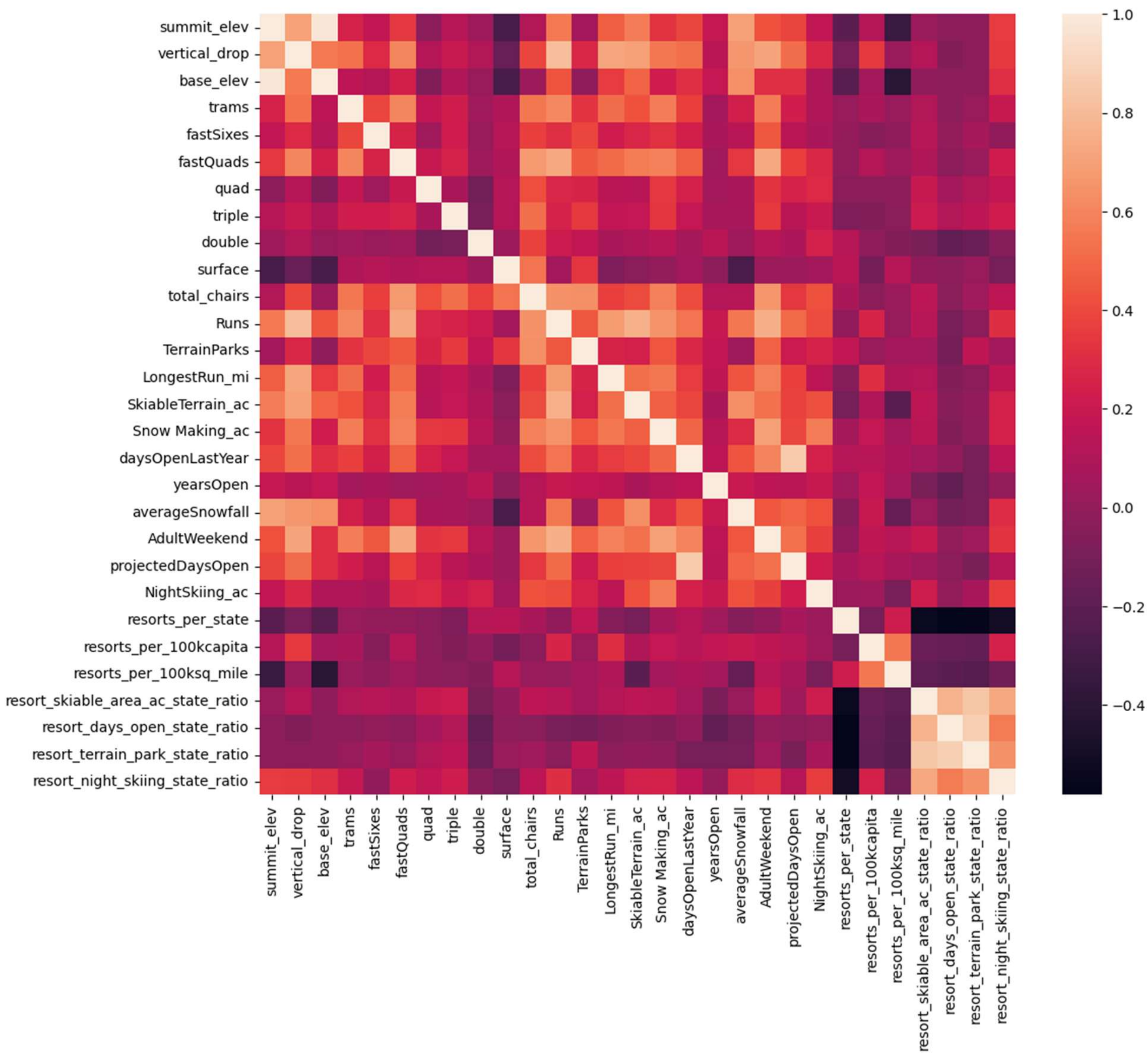
(Silverton Mountain) outlier and by removing the data for the 'yearsOpen' (2019) outlier. We also inspected states/regions differences, distributions of number of resorts by region/state and distribution of ticket prices by state. We derived state-wide summary statistics such as number of resorts, skiable area, days open, terrain parks, nighskiing area, population and area totals by state. We were left with 277 rows and 25 columns.

Exploratory data analysis

We used the PCA (Principal Component Analysis) technique to visualize data in lower dimensions and inspect how much variance in the price is explained by different features. First, as shown in the graph below, there was no pattern suggesting a relationship between state and ticket price. Therefore, we chose to build a model that considers all states together.



The results of our analysis are shown in the figure below:



We concluded that the features with the highest correlation to our target feature ‘AdultWeekend’ ticket price are: 'fastQuads', 'Runs', 'Snow Making_ac', 'total_chairs' and 'vertical_drop'.

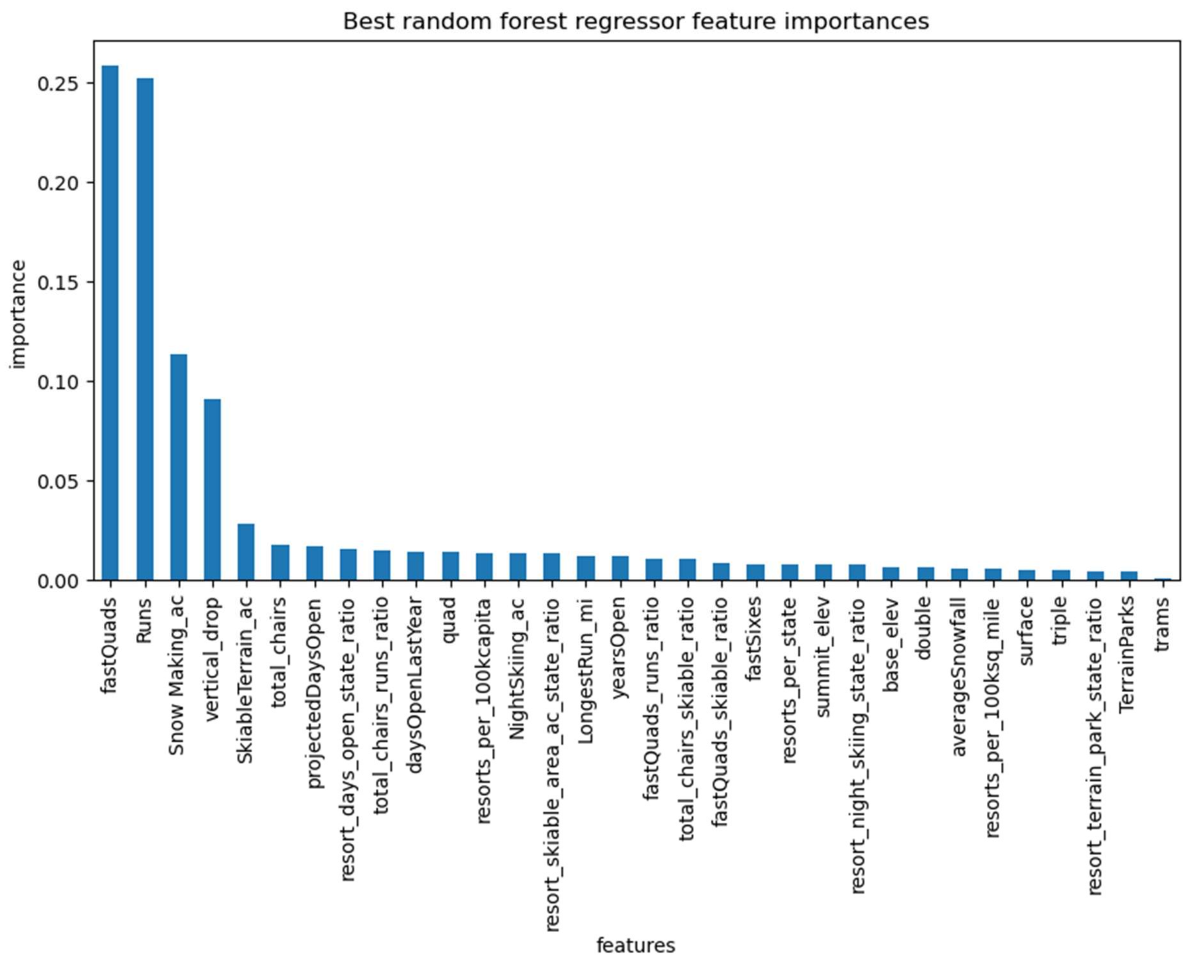
Model preprocessing with feature engineering

We split our data into a train/test split. We started by checking how good the mean average price is as a predictor using r-squared, mean absolute error, mean squared error and we expected to be off by around $19. Next, we imputed missing feature values for both train and test splits with predicted values using medians then means. This linear regression model led us to expect an estimated ticket price around $9 off the real price. This was an improvement from guessing using just the average.

Algorithms used to build the model with evaluation metric

Next, we tested a linear regression pipeline model based on sklearn's SimpleImputer function. The results were similar to the previous model, and we were still off by around $9 from the real price. Next, we refined our model with a new pipeline by selecting a different number k of features using sklearn's SelectKBest function. The default number of k=10 gave us worse results. Next, we selected the best value of k using sklearn's GridSearchCV function which gave us k=8 (vertical_drop, Snow Making_ac, total_chairs, fastQuads, Runs, LongestRun_mi, trams, SkiableTerrain_ac).

Next, we tried a random forest model. Its estimated performance via cross-validation was better than our previous linear regression model. We further improved our random forest model with hyperparameter search using GridSearchCV. The plot below shows each feature and its importance in our model. The dominant four features are fastQuands, Runs, Snow Making_ac and vertical_drop.

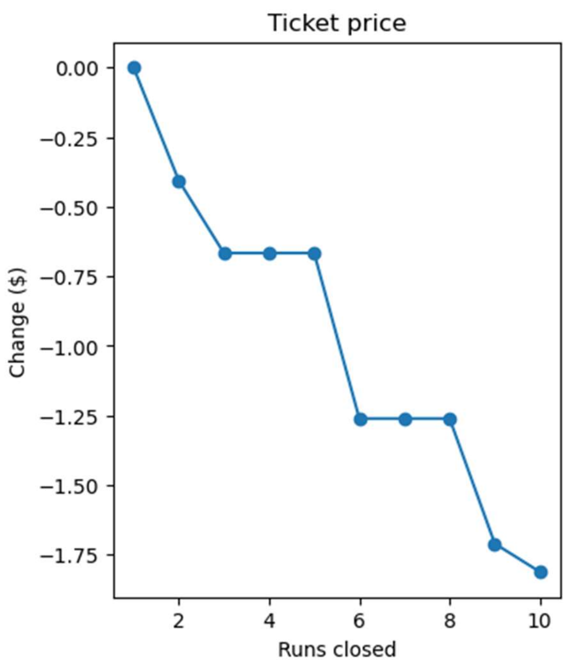


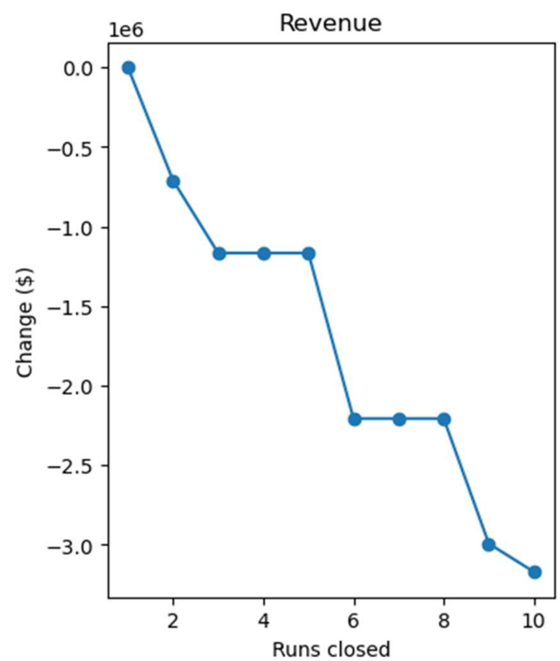
Winning model and scenario modelling

We have decided to choose our random forest model as it has a lower cross-validation mean absolute error which is consistent with the better estimated performance via cross-validation.

We modeled four different scenarios:

Scenario 1: Permanently closing down up to 10 of the least used runs. Permanently closing down 1 used run does not make any difference in ticket price support. Closing 2 runs reduces support for ticket prices. Closing 3, 4, or 5 reduces further ticket price support by the same amount. Closing 6, 7, or 8 reduces further ticket price support by the same amount. The graphs below show these results.





Scenario 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 prices by $1.99 and revenue by $3,474,638.

Scenario 3: Same as scenario 2 but adding 2 acres of snow making cover. This scenario increases support for ticket price and revenue by the same amount as scenario 2 ($1.99 and $3,474,638 respectively). Therefore adding 2 acres of snow making cover is not justified as scenario 2 is a better option.

Scenario 4: Increase the longest run by 0.2 miles to boast 3.5 miles length, requiring additional snow making coverage of 4 acres. There is no difference in ticket price support for this scenario. Therefore, an additional snow making coverage of 4 acres is not justified.

Pricing recommendation

Big Mountain Resort currently charges $81. Our model suggests an average ticket price of $95.87 with an expected mean absolute error of $10.39.

Conclusion

Scenarios 1 and 2 are recommended for further consideration. The business should close the least used run as it does not make any difference in ticket price support and then try to test for 2, 5, 8, 9 and 10 run closures successively if it is justified to do so. As for Scenario 2, the business should make sure that the additional operating cost of the new chair lift per ticket is less than the increase in ticket price support.

Future scope of work

Some information that would be useful is the cost related to maintaining each run and the cost related to the increase in vertical drop. We could incorporate this model into an easier interface or an app that business analysts could use to predict prices in different scenarios with parameters of their choosing.