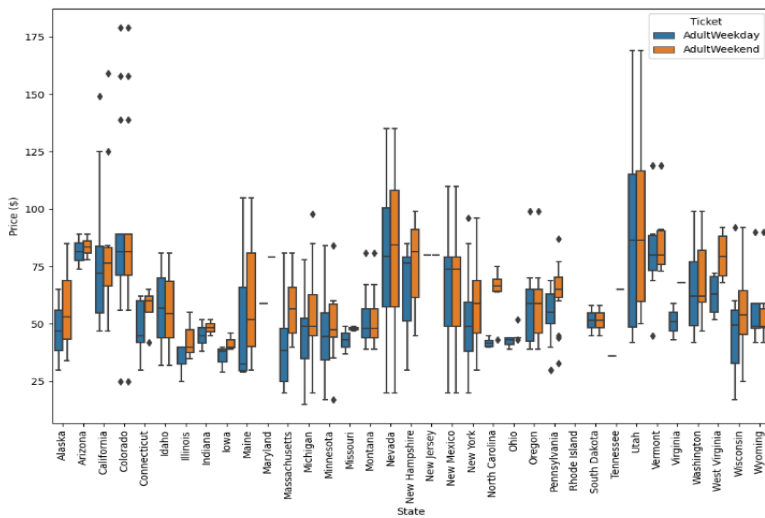


Guided Capstone Project Report

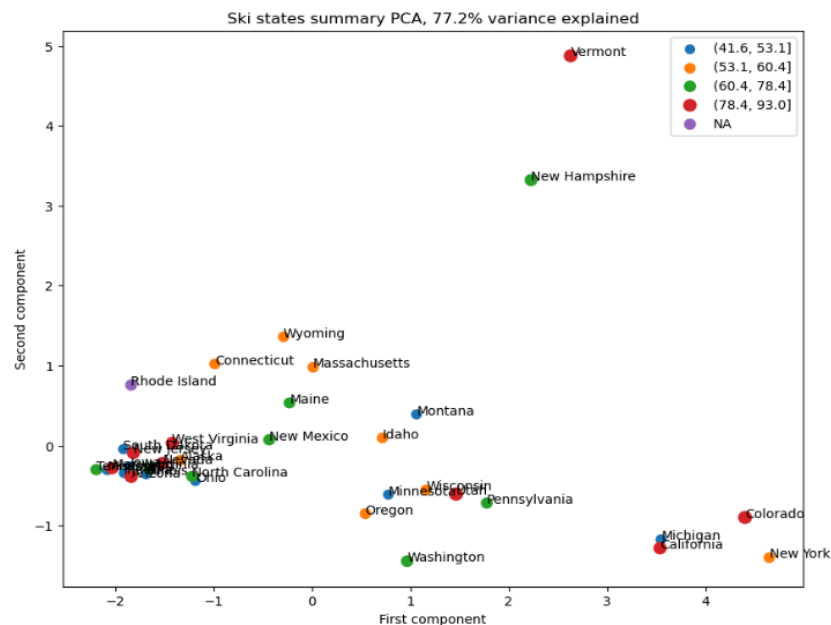
Big Mountain Resort in Montana charges higher than average prices compared to other resorts in its market, but does not have enough data about which facilities contribute to this premium cost. The resort offers many amenities, spectacular views, and is accessible to all skill levels. They would like to know the relative importance of the facilities offered to inform decisions about the future of the facilities offered by the resort.

Ski data for 330 ski resorts across America were analyzed, with 14% of resorts missing both weekday and weekend ticket price data. These 14% were removed from analysis. Weekend and Weekday ticket prices were combined. Data was analyzed for duplicate values, missing values, and outliers. Incorrectly high values for Skiable Terrain for Silverton Mountain were replaced with values from their website. 'fastEights' was removed entirely due to missing and incorrect data. Final dataset had 277 rows.

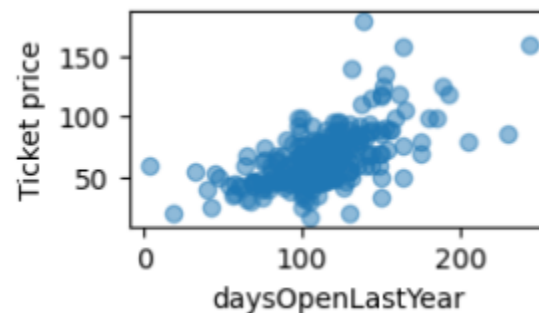
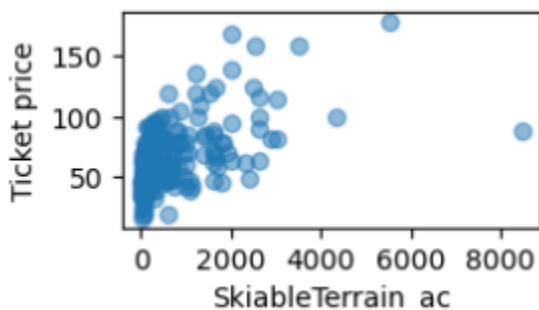
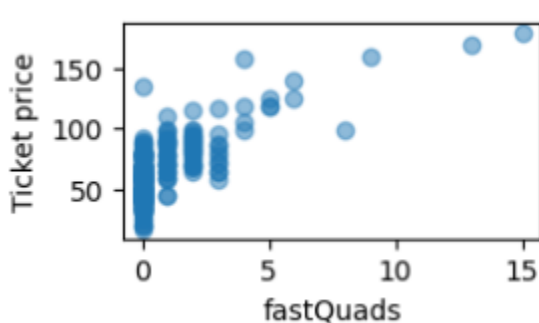
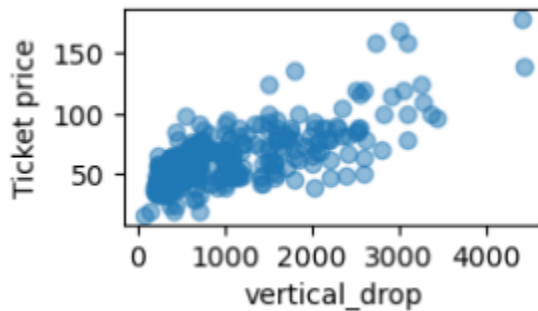
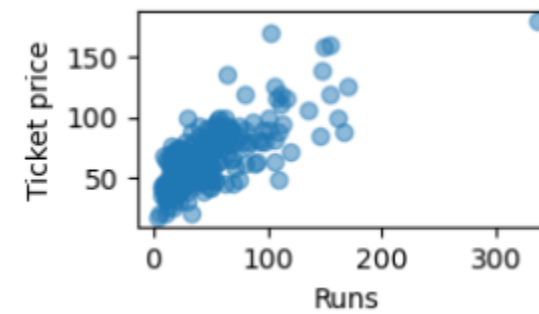
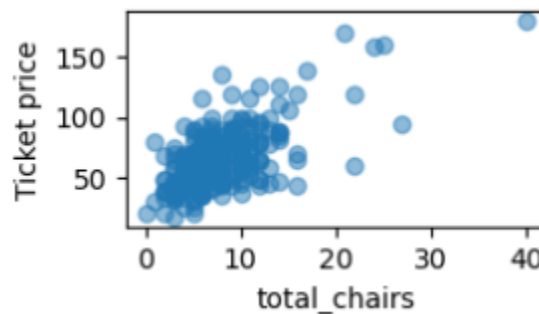
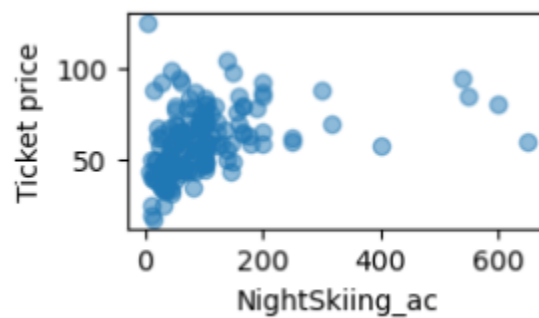
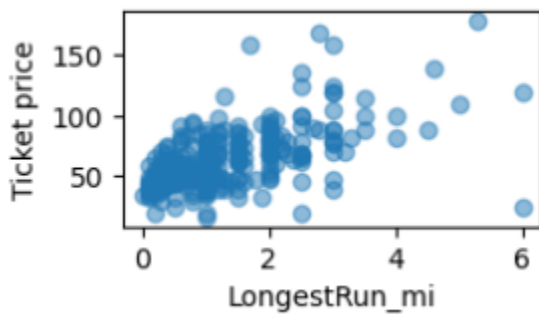


State-wide summary statistics were generated. Population of each state was added to the dataset. Montana showed much less variability in price than other states.

Exploratory data analysis on each state's resort summary was completed. New features, ratios comparing the number of resorts to state population and state size, were calculated and integrated into the dataset. Once the data was scaled and fit using PCA, the transformation showed that 75% of variance was from two components. The variance was plotted along with the state ticket price quartile. Not much of a pattern shown between state variance and ticket price variance, but does contribute to the understanding of the ski resort market.



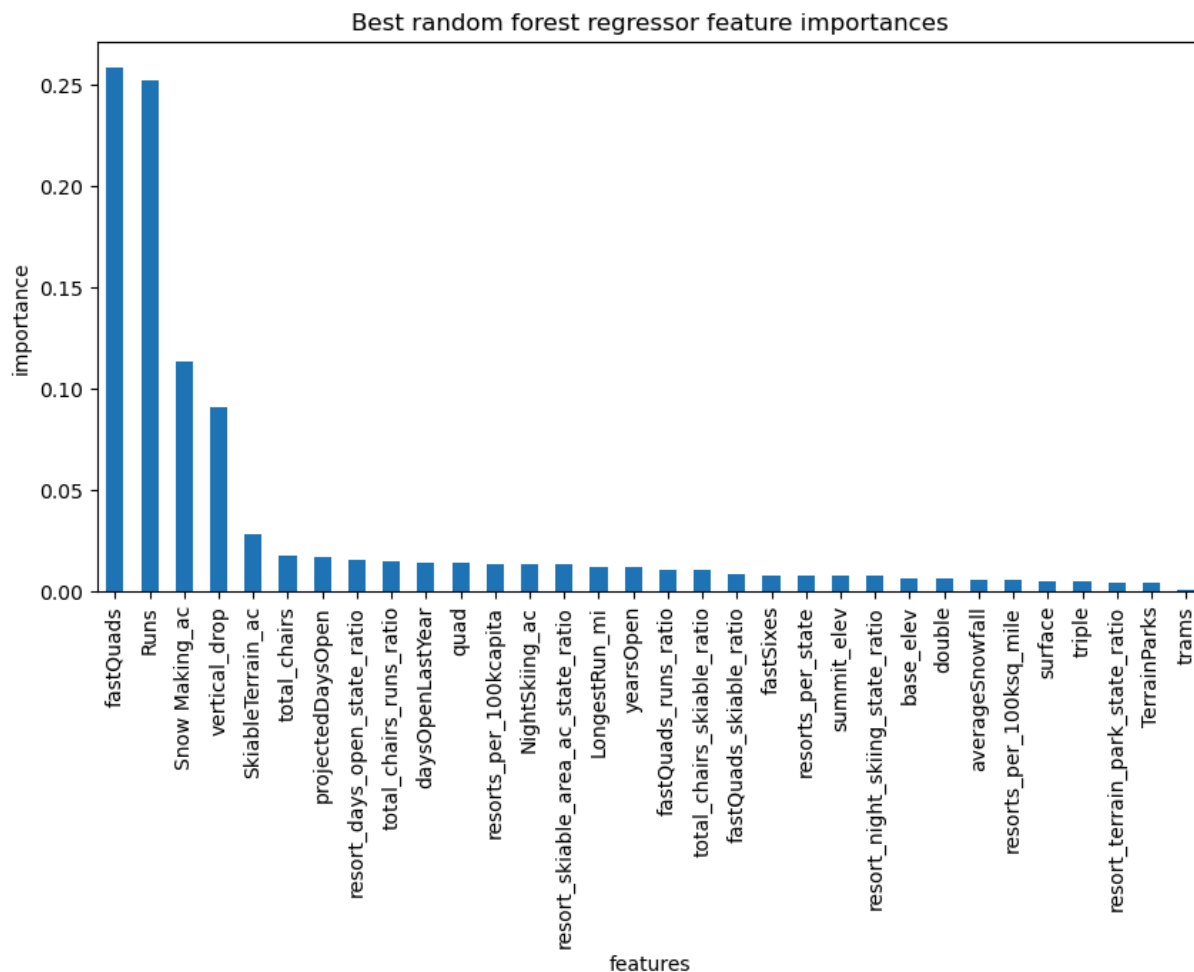
More 'resort-to-state' ratios were added into the data: 'resort_night_skiing_state_ratio', 'resort_terrain_park_state_ratio', 'resort_days_open_state_ratio', 'resort_skiable_area_ac_state_ratio'. These ratios, along with previous components were incorporated into scatter plots against ticket price. The components 'fastQuads', 'projectedDaysOpen', 'LongestRun_mi', 'Snow_Making_ac', 'total_chairs', 'vertical_drop', 'night_skiing_ac' and 'Runs' showed very positive correlation with 'Average Ticket Price,' our target feature. Amount of snow cover seems to be valued more than the amount of skiable terrain.



Finally, a linear regression was tried, experimenting with the number of folds in a cross-validation set and imputing missing values as median and mean. Hyperparameter tuning showed the ideal number of folds as 8, imputing missing values with median over mean. Each component was analyzed and vertical_drop was shown to be the biggest positive feature. Interestingly, SkiableTerrain_ac was shown to be a negative feature, which is not consistent with earlier exploratory data analysis.

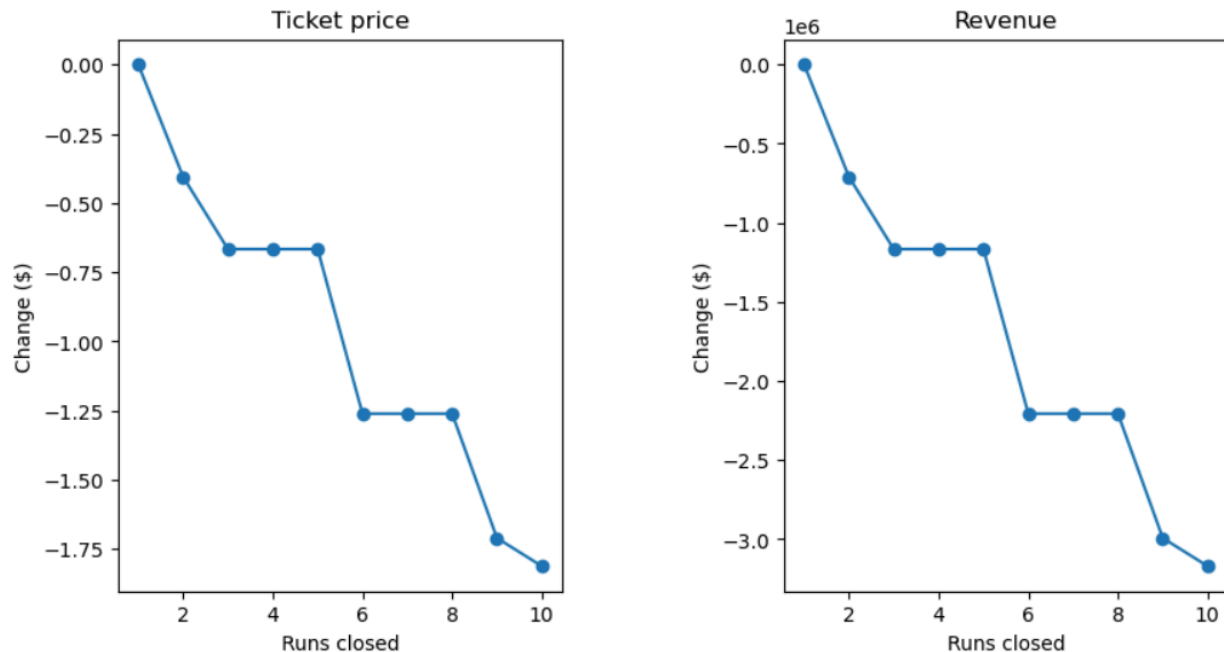
| | |
|-------------------|-----------|
| SkiableTerrain_ac | -5.249780 |
| trams | -4.142024 |
| LongestRun_mi | 0.181814 |
| Runs | 5.370555 |
| fastQuads | 5.745626 |
| total_chairs | 5.794156 |
| Snow Making_ac | 6.290074 |
| vertical_drop | 10.767857 |

Random Forest Regressor was tried next, imputing with median, not scaling the data as linear regression did, and using a number of estimators of 69. The feature importances were plotted below:



Random forest regressor was ultimately chosen because linear regression showed more overfitting and a MAE of 11.79 while random forest regressor showed a MAE of 9.53 and less variability.

After saving the model, Big Mountain's data was removed for testing purposes. Big Mountain currently charges \$81 per ticket. Modeling suggests that the market can support a \$14.87 increase in ticket price to \$95.87. MAE is \$10.39, showing there is room for at least a \$4.48 increase in price. Different business proposals were tested in the model, with revenue calculated based on an expected visitorship of 350,000. According to the model, closing 10 of the least popular runs would have the following effect on expected ticket price:



Closing 5 runs would incur a loss of \$245,000 ($0.70 * 350,000$) and closing 8 runs would incur a loss of \$3,500,000.

A second proposal to add a run, increase vertical drop, and install an additional chair lift was tested by the model. This would increase ticket cost \$1.99 which would lead to a revenue increase of \$3,474,638 per year. Costs of adding another chair lift would be \$1,540,000 per year.

I would recommend implementing the second proposal as well as closing 5 runs and raising ticket price \$7. This price increase is the sum of the \$1.99 suggested price increase from the second proposal plus the model suggested increase in price minus the mean absolute error (\$4.48) from Big Mountain analysis.

Fine tuning the hyperparameters of the model is still needed, including weighting the resorts most similar to Big Mountain. A user-friendly UI could be designed to change model parameters, estimate business proposals, and compare resorts.