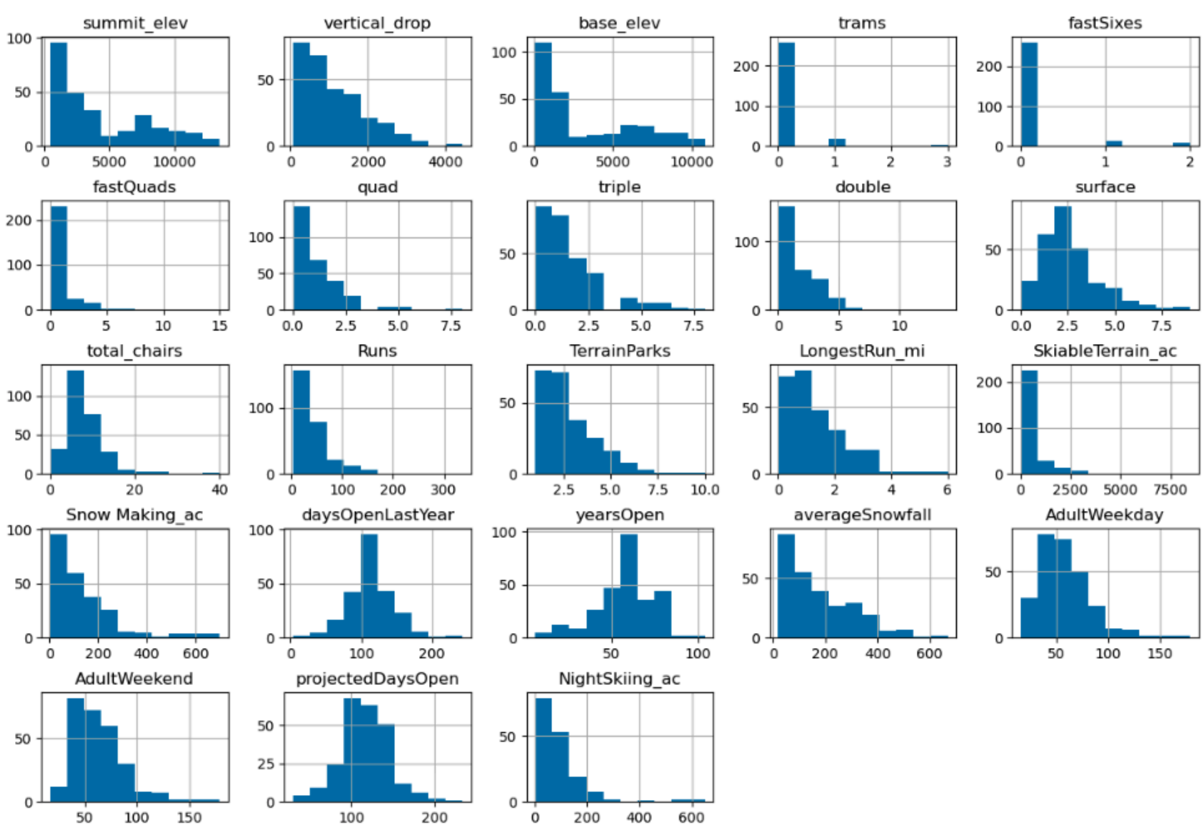
Guided Capstone Project Report

Problem Statement: How to select a better value for Big Mountain Resort's ticket price to either cut costs without undermining the ticket price or support an even higher ticket price in order to increase revenue for the resort by 20% within the next financial year.

Data Wrangling:

- Original Data: 330 rows X 27 columns. Big Mountain Resort was present.
- Additional cleaning: replaced incorrect values w/ correct ones, removed other unnecessary columns. Dropped records for 16% of resorts that were missing ticket prices.
- Target feature for ticket price prediction: Weekend Price. There are two separate ticket prices: Weekday Price and Weekend Price. Weekend Price is the target because it has less missing values than Weekday Price. Dropped rows with Weekday Price as a result.
- Investigated feature distributions, eliminated skewed ones, ended w/ these distributions:



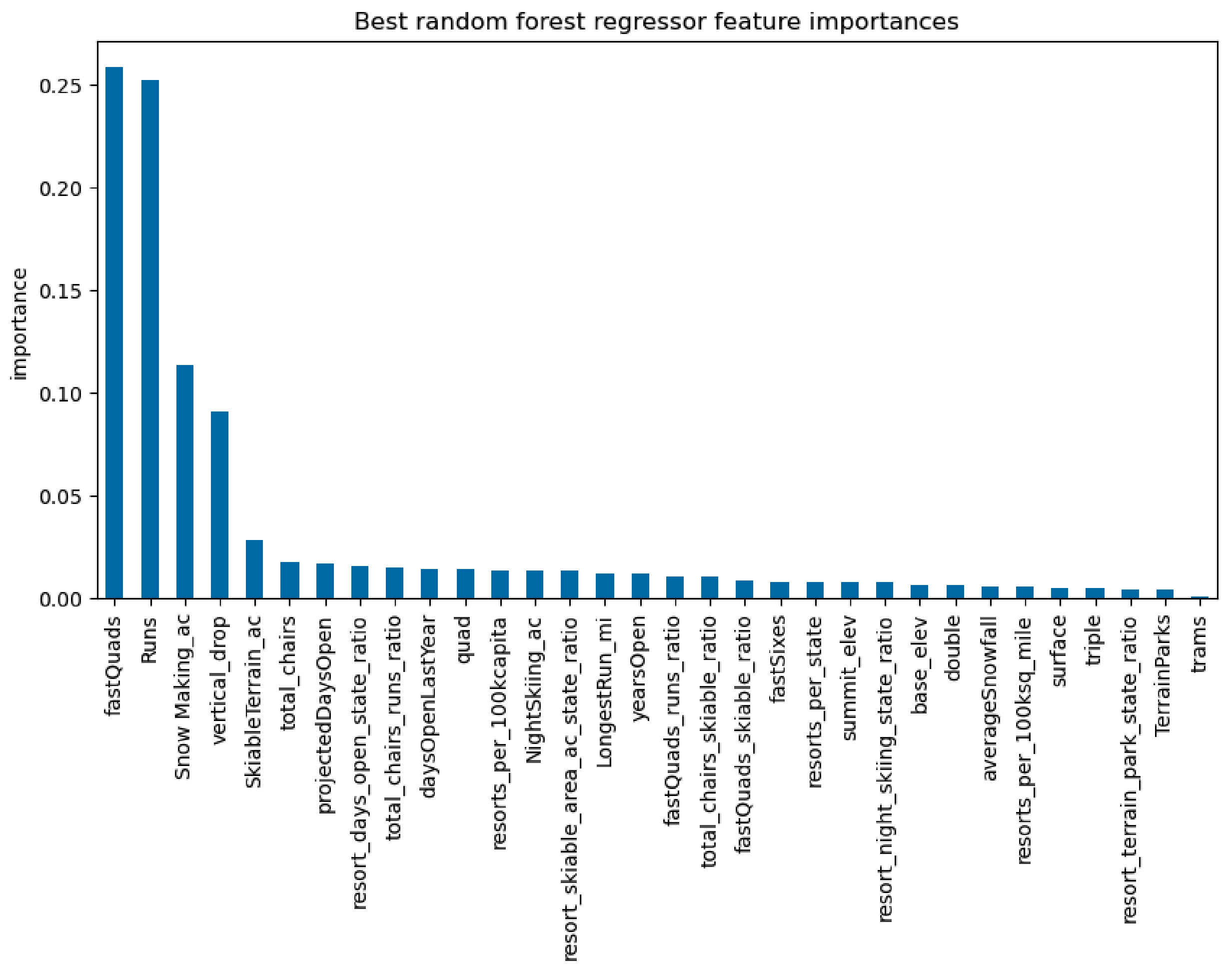
- Potential relevant features for predictive model:
    - TerrainParks

- SkiableTerrain_ac
- daysOpenLastYear
- NightSkiing_ac

- Rows and columns left in the dataset after cleaning: 277 rows X 25 columns.

Exploratory Data Analysis:

- There was a relationship between state and ticket price depending on each state's resort offerings.
- States differ in their # of resorts, skiable area, and night skiing, as well as resorts per population and resorts per area. Each of these is related to ticket price.
- We introduced multicollinearity with the new ratio features: resorts_per_100k_capita, resorts_per_100k_sq_mi, plus resort to state ratios for skiable area, days open, terrain park, and night skiing. They are negatively correlated with the number of resorts in each state.
- The features with the strongest correlation to ticket price were Snow Making_ac, Runs, FastQuads, Total Chairs, and Vertical Drop.



Pre-processing and Training Data:

- Built a simple linear regression model. Found this explains over 80% of the variance on the train set and over 70% on the test set.
    - Performance Estimate from Cross-Validation: R2 = .6327. Training split R2 = 0.7924096060483825, test split R2 = 0.637619997317079 (test split value was consistent w/ estimate from CV).
- Built a Random Forest Regressor; best pre-processing steps were imputing with the median which helped, but scaling the features didn't.
    - Performance Estimate from Cross-Validation: (0.7097946872278385, 0.0645871948748804) -- performance on test set was consistent with this.
- Final Model Selected: Random Forest model b/c it has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability.

Modeling:

- Big Mountain Resort currently charges $81.00; but the modelling suggests a ticket price of $95.87 could be supported in the marketplace.
- Even with the expected mean absolute error of $10.39, this suggests there is room for an increase.
- The additional operating cost of a new chair lift per ticket (assuming each visitor on average buys 5 day tickets) is $1.99.
- For further improvements, consider closing at least one run. The model says closing one run makes no difference to the ticket price. 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, but 6 or more closures leads to a large drop.
- If you were to test this, start with closing one run (knowing it won't have a negative impact), and then close a second one to see how quickly the revenue is affected. Wait until you have enough data before closing a third or more.

Conclusion and Future Scope:

- The business could make use of the model by experimenting with different scenarios of feature tweaking, for example increasing or decreasing the number of closed runs, and/or the number of snow making coverage, and/or other feature tweaks to see what the predicted ticket price and thus revenue impact would be. This would allow them data-driven insights before making significant business decisions.