Problem Statement

How can Big Mountain ski resort maximize its revenue for the next year, by implementing a new pricing & investment model starting next season, through selecting a new value for their ticket price, and/or through altering their upcoming investment strategy?

Data Wrangling

Source file: ski_resort_data with **330 rows for 27 columns**. This dataset included information on Big Mountain Ski Resort, with no missing values for the resort.

This dataset was merged with an external dataset containing **state size & population data** sourced from Wikipedia.

Dropped columns:

- 'fastEight': half the values in this column are missing and the rest are the value zero.
- 'AdultWeekday': 'AdultWeekend' is selected to be the target feature, so information on weekday pricing is not required for modeling.

Dropped rows:

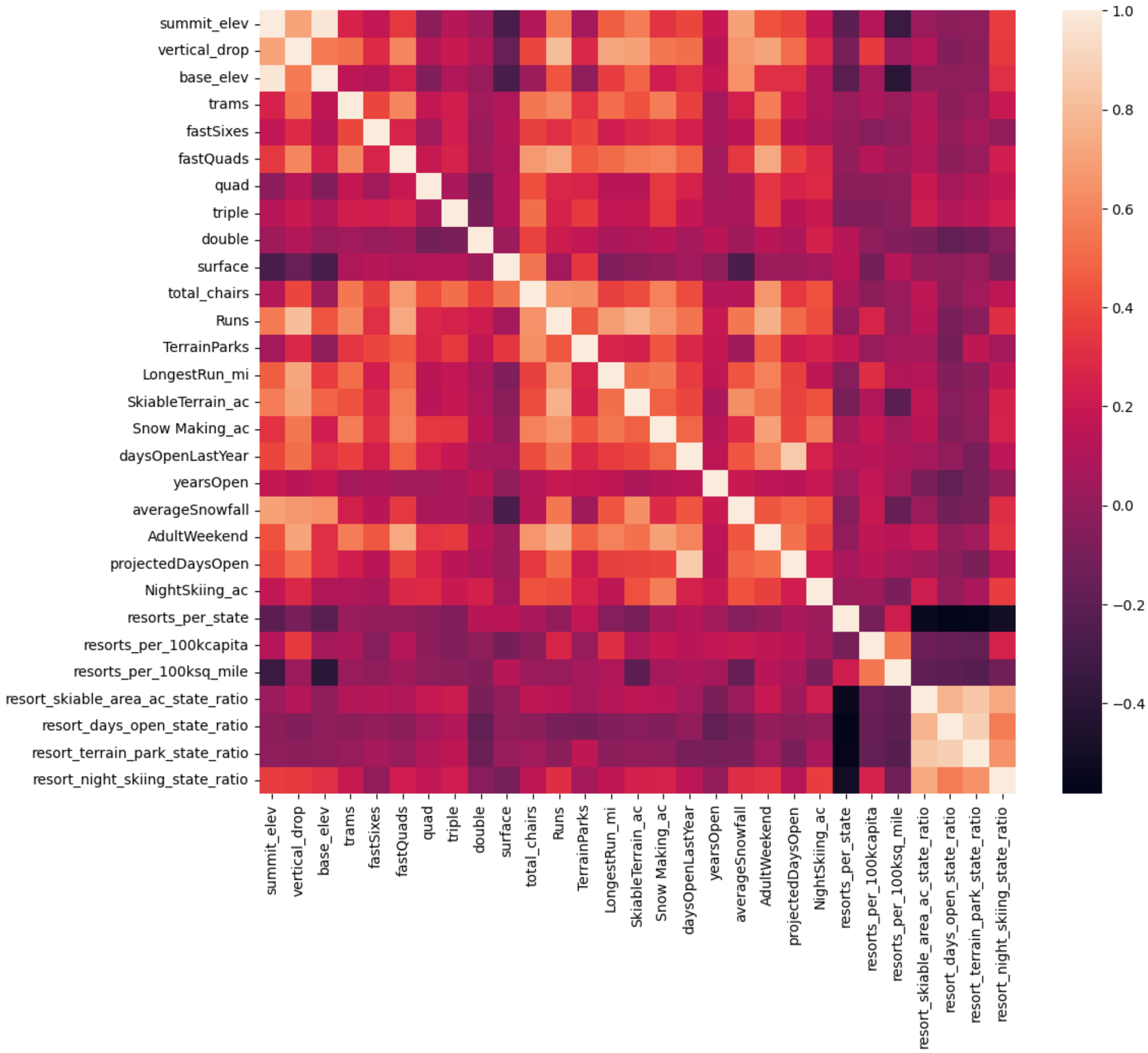
- Rows having missing values in the target feature 'AdultWeekend' were dropped.
- In the column 'yearsOpen', 1 value was incorrectly entered as 2019. Since a reasonable assumption could not be made for replacing the value, the entire row was dropped.

Post Data Wrangling, final dataset contained **277 rows** and **25 columns**.

Exploratory data Analysis

Target Feature to predict the ticket price was selected to be **'AdultWeekend'** for the following reasons:

- For the state of Montana, where the resort of interest is located, AdultWeekday values and AdultWeekend values were equal for all resorts.
- AdultWeekend had less missing values compared to AdultWeekday



Heatmap of correlation between features

Features which seemed important for modeling, after observation of collinearity with the target feature (**AdultWeekend**):

- ‘vertical_drop’, ‘fastQuads’, ‘Runs’, ‘total_chairs’, Snow Making_ac’, ‘LongestRun_mi’, ‘trams’, SkiableTerrain_Ac’, ‘resort_night_skiing_state_ratio’

Features showing high collinearity with each other, which we were wary of while modeling to avoid redundancy:

- Positive correlations:
  - ‘summit_elev’ with ‘base_elev’
  - ‘daysOpenLastYear’ with ‘projectedDaysOpen’

- Negative correlations:
    - ‘surface’ with ‘summit_elev’, ‘base_elev’, ‘averageSnowfall’
    - ‘resorts_per_state’ with ‘resort_skiable_area_ac_state_ratio’, ‘resort_days_open_state_ratio’, ‘resort_terrain_park_state_ratio’, ‘resort_night_skiing_state_ratio’

Other conclusions:

- After doing Principal Component Analysis on the state data, no pattern which could demonstrate a strong relationship between state label and ticket price stood out. In the process of building the pricing model, all states were considered equally.

Model Preprocessing with feature engineering

Engineered Features:

- To measure resort density:
    7. **‘resorts_per_100kcapita’**: ratio of resorts per state to state population (replaced ‘state_population’)
    8. **‘resorts_per_100ksq_mile’**: ratio of resorts per state to state area (replaced ‘state_area_sq_miles’)
- To put each resort within the context of its state:
    7. **‘resort_skiable_area_ac_state_ratio’**: ratio of resort skiable area to total state skiable area (replaced ‘state_total_skiable_area_ac’)
    8. **'resort_days_open_state_ratio’**: ratio of resort days open to total state days open (replaced ‘state_total_days_open’)
    9. **‘resort_terrain_park_state_ratio’**: ratio of resort terrain park count to total state terrain park count (replaced ‘state_total_terrain_parks’)
    10. **‘resort_night_skiing_state_ratio’**: ratio of resort night skiing area to total state night skiing area (replaced ‘state_total_nightskiing_ac’)
- To measure ease of transporting people in a resort:
    7. **‘total_chairs_runs_ratio’**: ratio of total chairs to number of runs
    8. **‘total_chairs_skiable_ratio’**: ratio of total chairs to skiable terrain area
    9. **‘fastQuads_runs_ratio’**: ratio of number of fast quads to number of runs
    10. **‘fastQuads_skiable_ratio’**: ratio of number of fast quads to skiable terrain area

Algorithms used to build the model with evaluation metric

Metrics for the **Baseline model** - using the average (mean) price as a predictor:

- R-squared: -0.0031235200417913944
- MAE: 19.136142081278486
- MSA: 581.4365441953483

Best **Linear Regression** model results:

- Accuracy of 63.3%, with a variability of 44% to 82%.

- MAE:
    1. Training data: 10.5
    2. Test data: 11.8

Best **Random Forest** model results:

- Accuracy of 69.8%, with a variability of 58% to 84%.
- MAE:
    1. Training data: 9.65
    2. Test data: 9.54

Winning model and scenario modeling

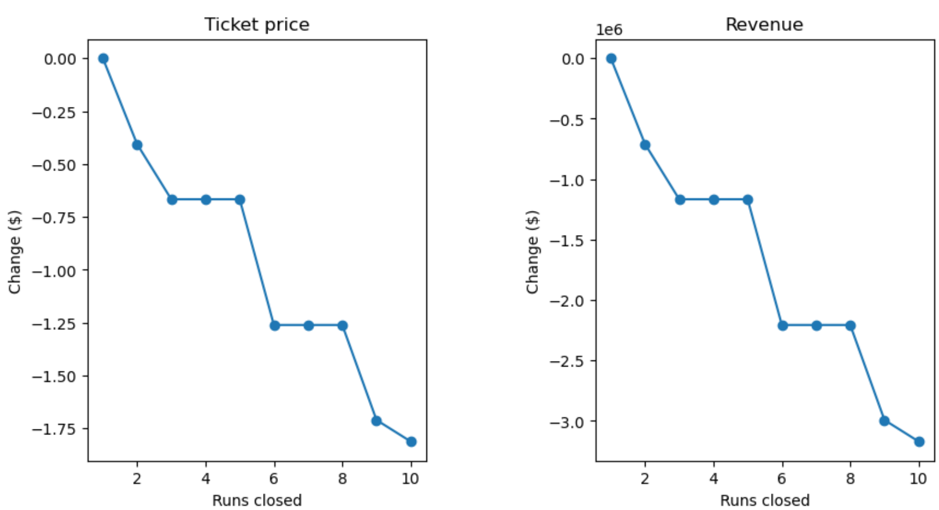
Model Selection:

- Random Forest displayed a **lower cross-validation MAE** compared to that of Linear Regression.
- Random Forest's performance on the test set is **more consistent** with the cross-validation estimates.
- Random Forest also exhibits **less variability** (58-84%) compared to that of Linear Regression (44-82%).

Hence, **Random Forest is chosen as the model to be used going forwards.**

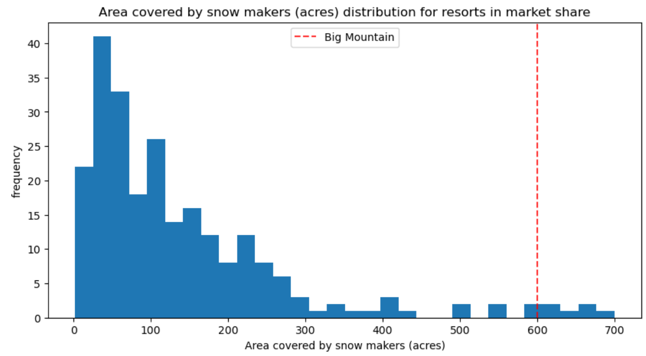
Scenario Modeling:

1. **Closing Runs:**
    - According to the model, **closing 1 run will make no difference on the revenue**.
    - Closing 2 and 3 runs successively will lead to a **reduction** in support for the charged ticket price, and hence, the revenue.
    - If business decides on closing 3 runs, they can go ahead and close down 4 or 5 runs, as this will **not lead to any further loss** in revenue.
    - Closing **more than 5 runs** will lead to a **significant drop in revenue**.

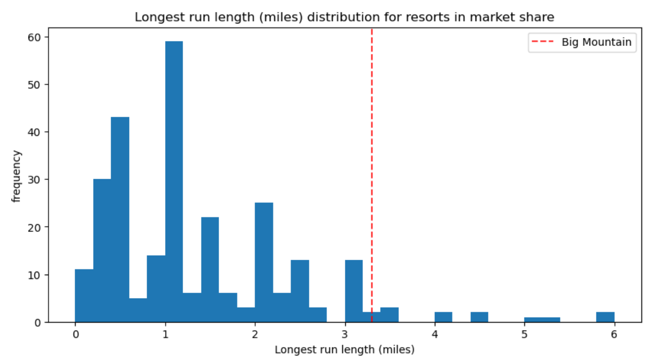


2. **Increasing the vertical drop by adding a run and installing an additional chair lift:**
    - An **increase of 150 ft** in the vertical drop, after adding a run and a chair lift, increases support for the ticket price by **$1.99**.
    - Over the season, this change could lead to an increase of about **$3,475,000** in the total revenue.

3. **Increasing snow making cover, along with changes suggested in the previous point:**
    - An **increase of 2 acres** in the snow coverage is **too small** to make any difference in the projected revenue.
    - Big Mountain is already very high up the league table of snowmaking area. Investing in more coverage will not significantly affect its position among its competitors.



4. **Increasing longest run to 3.5 miles, with an increase in the snow coverage of 4 acres:**
    - An **increase of 0.2 miles** in the longest runs will **not lead to an increase in the projected revenue.**
    - Big Mountain already has one of the longest runs in the country. Resorts with longer runs are rare.

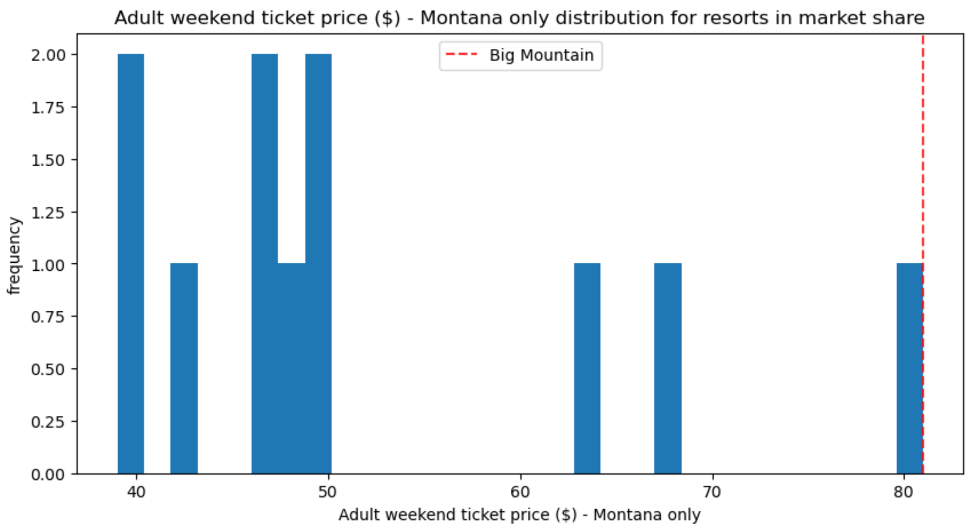


Pricing recommendation

- Current ticket price charged by Big Mountain: **$81.00**
- Ticket price suggested by modeling: **$95.87**, with a mean absolute error of **$10.39**. In other words, modeled ticket price ranges between **$85.48** and **$106.26**.

Conclusion

- Modeling clearly shows there is room for increase in the ticket price charged by Big Mountain, when being compared with prices set by the rest of the market, based on the facilities provided.
- To cover the additional operating costs of **$1,540,000** this season, caused by the installation of the new chair, an increase of at least **$8.80** in the ticket price is required (based on 350,000 expected visitors buying 5 day tickets, on average).
- Changing the ticket price based on calculations made in the previous point, would raise the ticket price to about **$90**. This figure lies in the range predicted by our modeling.



- Big Mountain is already charging the highest ticket price in the state of Montana. This is a testament to Big Mountain not compromising on the quality and quantity of the facilities provided by them. Therefore, changing the ticket price to a value, which accurately reflects that this is the best resort in the state, makes business sense.

Future scope of work

- More information such as **total operating costs, number of visitors per year, number of tickets sold per day** would definitely assist in making sharper predictions.
- Model and packages information is saved in the **pre-processing file** (04_preprocessing_and_training), for re-use purposes.