Guided Capstone Executive Presentation

Maritza Sazo
Data Scientist

Problem Identification - Context

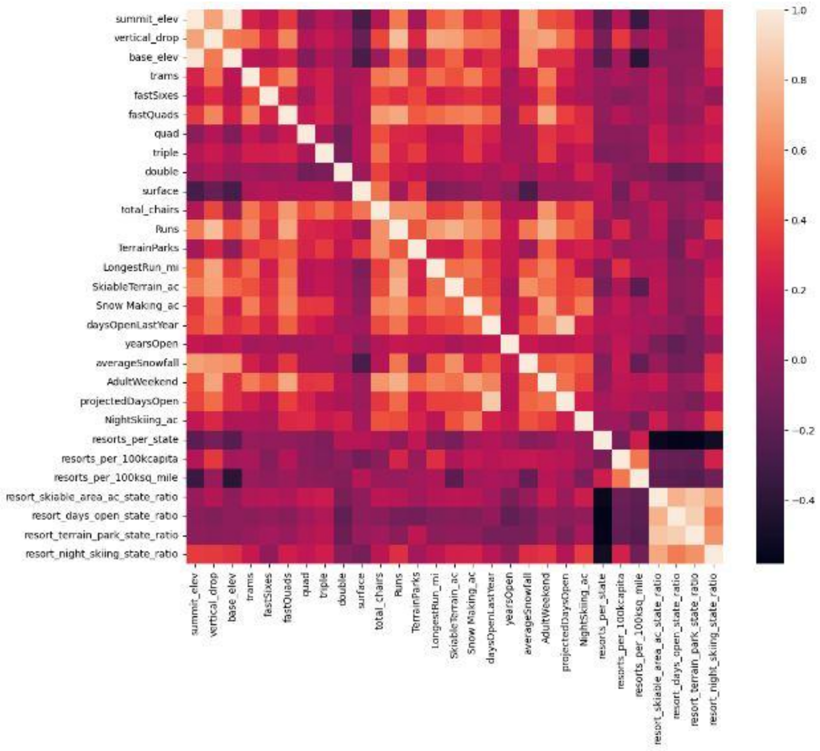
- Big Mountain Resort (Montana)
    - Seeking to reevaluate ticket pricing strategy and optimize facilities
        - Current strategy: premium pricing above average of resorts in market segment
        - Suspicion that facility usage is not optimized
        - No metric to gauge which facilities are more important/valued
    - Current adult ticket price is $81 on weekdays and weekends
    - New chairlift has added $1.54 million in operating costs
- Provided dataset includes data from ski resorts across the US with mostly numerical data (incl. Summit elevation, vertical drop, number of runs at resort, adult weekday ticket price, adult weekend ticket price) and some categorical data (name of resort, region, state)

- **Can we make a predictive model for ticket prices that reflects the facilities and features the resorts have to offer, and use it to determine what Big Mountain should charge for their facilities?**
- **What type of facilities optimizations can Big Mountain perform to justify an increase in ticket price?**

Recommendations and Key Findings

- Increasing the vertical drop by adding a run to a point 150 ft lower down, combined with the additional chair lift, with ~ 2 acres of snow making coverage justifies a ticket price increase of $1.99 to $82.99
- This corresponds to a revenue increase of $3,474,638, surpassing the additional operating costs.
- Overall, our model supports a $95.87 ticket for Big Mountain Resort, given its facilities and features
- Key findings
  - State-level analysis of the data generated a lot of hypotheses and questions regarding the distribution of different features across ski resorts in different states, and even within the same state
  - However, ticket prices were not found to have any distinct relationship with state.
  - The four features with the highest positive correlation to adult weekend ticket price (i.e., that provided the most value) were number of fast quads, number of runs, vertical drop, and total number of chairlifts.
  - We ran two models, a linear regression and a random forest, and both models confirmed the importance of these features, as well as snow making acreage, in determining ticket price.
  - Ultimately, our random forest model showed less variability than linear regression and was used to predict Big Mountain's target ticket price.

Modeling Results and Analysis



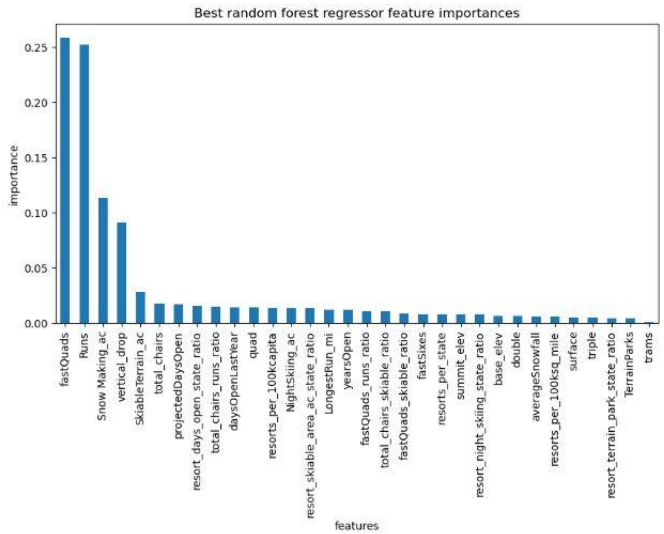
- Before modeling,
  - Data wrangling cleaned up missing or inaccurate values, helped visualize the distributions of features across resorts and across states, and allowed us to join state population, state size and resort density data to our ski resort dataset.
  - Most importantly, it helped us determine that our variable of interest in our model would be **Adult Weekend Price**.
  - Exploratory data analysis helped identify the features most highly positively correlated with adult weekend price - fast quads, number of runs, number of chairlifts, vertical drop,
  - This gave a preliminary idea of which features were most valuable in terms of justifying a higher ticket price.

Modeling Results and Analysis

- Linear regression model
  - Our initial model used linear regression to predict ticket price from the other numerical data values in the dataset.
  - The statistics used to assess model performance were R^2, mean absolute error, and mean squared error.
    - These metrics helped assess variance of our model and accuracy of its predictive capabilities.
  - Training and testing of the model determined that 8 features were needed to have the highest performance.
    - At this stage, we assessed performance with cross-validation.
  - The most significant features in our linear regression model were vertical drop, snow making acreage, number of chairlifts, fast quads, number of runs, longest run, number of trams, and skiable terrain acreage.
    - The last two features listed above had negative correlations with ticket price.

Modeling Results and Analysis

- Random forest model
  - Our secondary model was a random forest model that used the median to impute null values, non-scaled data and 5-fold cross-validation to assess performance.
- The four most significant features for this model were fast quads, number of runs, snow making acreage, and vertical drop.
  - These values showed overlap with our linear regression model and our correlation heatmap from EDA.
- To the right, you can see a visualization of the importance of each feature in our best performing random forest model.



Modeling Results and Analysis

- To select our final model, performance assessment values for both models were compared to each other.
- Random forest model was chosen due to more accurate predictive capabilities (lower MAE) and less variability.
  - Predictions using this model are closer to actual values by more than $1
- At this stage, more data collection is not necessary to further train the current model; however, more data could help better define the relationships between seemingly unrelated variables that could be affecting ticket pricing.

Summary and conclusion

- Our model predicted an overall ticket price of $95.87 for Big Mountain Resort, based on the prices and facilities/features of other ski resorts in our market segment.
- Modeling different scenarios of changing facilities allowed us to predict how much we could increase the ticket price based on the updates.
  - Depending on the importance of the feature in the model hierarchy, we can see how certain changes affect ticket prices more, and to what extent.
  - Closing runs supports a decrease in ticket price, starting at 2 runs and rapidly decreasing after 6 runs. One closed run has no effect, which could benefit the resort by decreasing operating costs with no lost ticket revenue.
  - Increasing snow acreage by small amounts does not affect ticket prices, although higher snow making acreage is overall a strong predictive feature in our models. Small additions of snow making acreage in support of other facilities optimizations may also help justify increasing ticket prices further.
  - According to our model, lengthening the longest run will not have any effect on ticket prices.
- We can justify a $1.99 ticket price increase with our current model by leveraging the newly added chairlift and adding a run that increases the vertical drop by 150 ft.
- **The model can be used to make further predictions on how changing certain features will affect ticket price.**