**Optimizing Big Mountain Resort’s Pricing Strategy through**

**Predictive Modeling**

**Executive Summary:**

**1. Problem Statement:** How to maximize capitalizing on Big Mountain resort's facilities in upcoming season so that cover $1,540,000 operating costs, increase the income and decide on company's future investment strategy.

**2. Data Wrangling:** Our main goal was to pre-process our data to improve its quality and efficiency. As a result, we removed individual columns and rows with empty values (or those of no practical value) and also replaced the distorted records with valid data from the Internet. We solved the problem of having two potential target variables by dropping the attribute with a large number of empty entries. We expanded our analytical capabilities by collecting additional publicly available information from the Internet about the population and size of each U.S. state. Apart from that we was able to catch some useful insights of our data which had important meaning for our price modelling.

**3. Exploratory Data Analysis (EDA) and Feature Engineering:** At this stage, we used PСA technology to delve into the analysis of data taken from the Internet and try to understand whether there are certain patterns and relationships between the ticket price and the State of location.

A graph of states with numbers and names

Description automatically generated with medium confidence

The visualization showed a spread of average prices across states without any patterns in the distribution (the location of the resort in the State will not be taken into account in the pricing model). At the same time, we identified the relevance of the studied data to our business model.

On this basis, we combined original data with information collected from the Internet, simultaneously using attribute engineering technologies. In particular, using simple arithmetic operations we obtained the following features:

* ratio of resort skiable area to total state skiable area;
* ratio of resort days open to total state days open;

- ratio of resort terrain park count to total state terrain park count;

- ratio of resort night skiing area to total state night skiing area.

As a result, we received a more informative data set with new useful attributes, which increased the efficiency of our analysis.

At the next stage of our analysis, we studied the correlation, which revealed multicollinearity between individual features, and also allowed us to draw preliminary conclusions about infrastructure elements that have a noticeable impact on the final ticket price.

A screenshot of a computer screen

Description automatically generated

In particular, we could cautiously conclude that resorts with greater snow cover, vertical drop, more Runs, list chairs and quad bikes, as well as night skiing opportunities, could charge higher ticket prices.

**4. Model Preprocessing and evaluation:** In the next step, we focused on building, testing, and evaluating our models using the following metrics: R2 (*the proportion of variance in the dependent variable that is predictable from the independent variable*), Mean absolute error (*the average of the absolute difference between the actual and predicted values*) and Mean Squared Error (*the average of the squared difference between the original and predicted values*).

We built Linear regression and Random Forest models, for each of which we tested

various combinations of hyperparameters, features and data preprocessing conditions using the tools of the sklearn library (*pipeline, GridSearchCV*). This allowed us to select the optimal set of parameters and identify the most useful features, thereby avoiding overtraining of models.

To ensure the validity of our model, we evaluated its performance using cross-validation technique, which guaranteed what testing was hold on unseen data.

**5. Winning Model and Scenario Modelling:** Based on cross-validation testing results, the Random Forest model showed the best performance.

MAE on the train split: 9.64 (*on average, the predictions of the model are about $9.64 away from the actual ticket price*).

The standard deviation of the MAE scores: 1.35.

MAE on the test split: 9.54 (*performance consistent with the cross-validation results*).

These indicators were achieved by the model using non-scaled data with a median strategy for imputing missing values. Four key attributes were identified that have the greatest influence on price prediction: snow cover, vertical drop, runs and quad bikes.

Using our model, we were able to predict that in an ideal market, our tickets would be priced at $95, which is 20% more than the current price.

Using our model, we tested all potential scenarios developed by the Big mountain resort. Scenario No. 2, which involves adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift was identified as the most optimal, since it would allow us to increase the ticket price by another $1.99 and increase revenue to $3474638.

**6. Pricing Recommendation:** According to our model, we can estimate the cost of tickets at $95, which is 20% more than the current price. Despite some doubts, we have certain arguments in favor of increasing ticket prices.

Big Mountain ranks high in almost all infrastructure facilities, which has a serious impact on the attractiveness of the resort. At the same time, the Big Mountain current prices are more than twice as low as the most expensive resorts in the country (81 vs 180).

If we trust the result of model our income will be around $24 500 000 at the end of the season. Thus, we will fully cover our operating costs associated with an additional chair lift in the amount of $1,540,000.

**7. Conclusion:** The model predicts the final value of price with certain caveats. But it provides valuable business analytics, as it allows you to take into account market conditions when forming a pricing policy. If we use the model correctly, we will be able to increase the company's profits by maximizing the use of our facilities.

**9. Future Scope of Work:** As we noted above our predicted price is based on the assumption of an ideal market and relevant prices for other resorts. Moreover, even at the stage of preprocessing and EDA we assumed a possible lack of data for the model (number of visitors each season). But an even more important condition for determining the correct pricing strategy is the availability of information about the operating costs of maintaining the facility.

The presence of this information would make it possible to improve the quality of the model’s work, and, perhaps, to revise scenarios for further development of pricing policy.