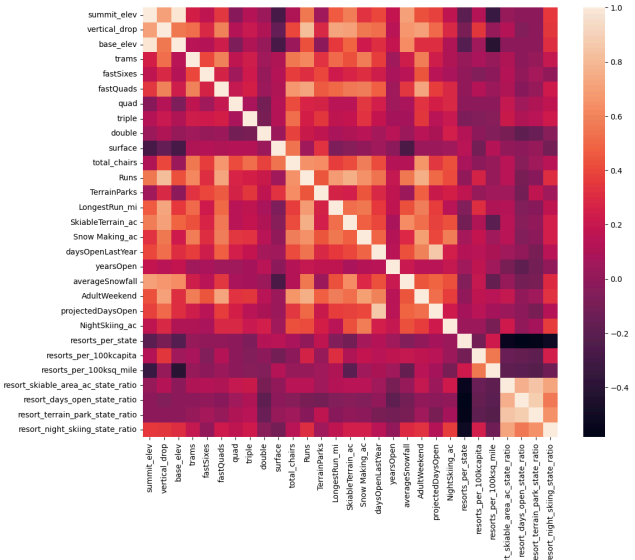
Guided Capstone Big Mountain Resort

Written By: Zachary Kaiser

The goal of this guided capstone project was to provide an introductory guided practice through the data science method, the goal pertaining to the actual data and workflow was to develop data driven pricing recommendations to Big Mountain resort. Throughout the process I had to work through each step of the data science method in a hands-on approach that allowed me to push my technical understanding and ability to interpret the data in a meaningful way.

The first step was identifying the problem: To find the best way to recommend pricing strategy adjustments in the most optimal way so that we could maximize revenue while still keeping it simple and very customer friendly. After identifying the problem and the stakeholders we had to take a look at the data. The raw dataset came with a lot of inconsistencies, null values, duplicates, and incorrect data types. We had to take an approach on cleaning that up. This included imputing missing values and transforming features like date and time into something more usable. I also encoded categorical variables and made sure everything was formatted correctly for the next stage of the data science method.

Exploratory data analysis (EDA) helped me understand the key drivers behind sales and pricing. Visualizations like histograms, box plots, and heatmaps were used to highlight patterns and correlations. One big insight was how product type and promotional offers significantly affected pricing success. The EDA also showed some interesting seasonality trends that could be useful for future scenario modeling. It was during this stage that the data really started to “tell a story.”



Once the key trends were found, I focused on preprocessing. I scaled numeric features using standardization and added interaction features based on the insights from EDA. For instance, combining discount rates with seasonality gave a better representation of customer behavior. This was an important step, I had to ensure that I could work with the data in a meaningful way,

if it were just raw data being shoved into an algorithm we would not expect great results. Bad data in = bad results out.

I tried out multiple algorithms including Random Forest, Gradient Boosting, and Linear Regression. Out of these, the Gradient Boosting model performed the best in terms of RMSE and R^2 score. I also conducted hyperparameter tuning using GridSearchCV to make sure I was squeezing out as much performance as possible. What made this part interesting was comparing different models based on both accuracy and real-world applicability. The best-performing model was then used for scenario modeling to see how changes in features (like price or promotion) could affect predicted outcomes.

Using the best model, which happened to be random forest, I ran different “what-if” scenarios to understand how various pricing decisions might play out. For example, I modeled what would happen to predicted sales if we dropped the price by 10% or removed a promotional discount. These simulations provided a foundation for building a dynamic pricing strategy rather than relying on fixed rules.

In conclusion, the model I built can be used not only for pricing predictions but also as a strategic tool to evaluate different business decisions. That said, there’s still room for improvement. Future work could involve adding more external data and exploring real-time pricing adjustments with a live dashboard based on some live related values from the resort. Overall, this project helped me better understand how data can drive smarter business choices, and it gave me a chance to apply theory in a real, impactful way.