Big Mountain Resort Report

The Big Mountain Resort’s price and investment strategy was investigated by using the data of 330 ski resorts of US.

The data sheet was a 330×27 data sheet with 27 columns. The Big Mountain Resort was on row of 151. 2 columns were removed. One is 'fastEight' , which has half of the data missing and the rest is value zero and contains almost no information. The other column removed is 'AdultWeekday'. This research will only focus on the ‘AdultWeekend’ price. 48 rows were removed because they had 2 price values missing. 4 rows were deleted because they had 'AdultWeekend' data missing. one resort reported unusually high open years of 2019, it is obviously wrong, and that row was removed. Finally, the cleaned version of the ski data was 277×25 sheet.

The PCA could not find a good relationship between the ski resort price and the name of the state they are located. It offered some justification for treating all states equally, and work towards building a pricing model that considers all states together, without treating any one particularly specially. The columns of ‘Name’, ‘state’ and ‘Region’ were dopped. On the other hand, 7 state summary features (resorts\_per\_state, state\_total\_skiable\_area\_ac, state\_total\_days\_open, state\_total\_terrain\_parks, state\_total\_nightskiing\_ac, resorts\_per\_100kcapita and resorts\_per\_100ksq\_mile ) and 4 ratios ( total\_chairs\_runs\_ratio, total\_chairs\_skiable\_ratio, fastQuads\_runs\_ratio and fastQuads\_skiable\_ratio) were added to the column. So it became a 277×33 data sheet. The row of Big Mountain Resort is not include in the modeling. So it is 276×33 for the modeling.

Among the 276 sets of data, 193 (70%) is for the model training and 83 (30%) is for the model testing. We tried to see the relationship between the price (AdultWeekend) and the other 32 numeric features.

If we just took the average of all the prices as the price prediction, the MAE (Mean Absolute Error) would be 17.9 for the training data set and 19.1 for the testing data set.

We used the median or mean to replace the missing data and linear regression method as the initial modeling way. The MAE was 8.5 for the training and 9.4 for the testing. The model showed the vertical drop is the biggest positive feature and the skiable terrain area is a negative feature. Finally the random forest regressor was used for the modeling, it gave MAE 9.5 for the testing. The top 4 dominant features are fastQuads, Runs, snow making area and vertical drop.

The model predicted the Mountain Resort price of $95.87, since the current price is $81, even with a MAE of $10, there is still some room for the price increase.

The features important for the price were vertical drop, Snow making\_area, , total chairs, fast Quads, Runs, Longest Run, trams and Skiable Terrain area. Closing 10 runs will decrease the price by $1.75 and cause a revenue loss of $3,000,000. Increasing the vertical drop by 150 feet will increase the price by $1.70 while adding a chair lift can lift price by $0.29.

This prediction model can be used as a price and investment decision guide.