Big Mountain Resort

Contents:

Introduction

Big Mountain Resort is located in northwestern Montana. Big Mountain Resort offers spectacular views of Glacier National Park and Flathead National Forest. The resort originally opened in 1947 with an annual snowfall of 333 inches and 3,000 acres of skier and rider accessible terrain. Big Mountain Resort offers access to 105 named trails and vast bowl and tree skiing. All these are serviced by 11 lifts, 2 T-bars and 1 magic carpet for novice skiers. The longest run is named Hellfire and is 3.3 miles in length. The base elevation is 4,464 feet, and the summit is 6,817 feet with a vertical drop of 2,353 feet. With a terrain rating of 12% beginner, 38% intermediate, 44% advanced and 6% expert, this mountain can accommodate skiers and riders of all levels and abilities.

Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000 this season. Every year about 350,000 people ski or snowboard at Big Mountain. This business profit margin is 9.2% and the investors would like to keep it there.

Problem Identification Overview

In this project , I would recommend the factors which will help to recover the additional operating cost and increase revenue by 9.2% in this season.

I have used K-means clustering technique and multiple linear regression model for best results

Data wrangling: Data cleaning

There are 330 rows and 27 columns.

There are 13 columns out of 27 with missing data . Most of these are missing less than 10% of the data. Since there is less percentage of missing data , it is not dropped and imputed by 0 or mean as per data. There are no duplicate rows

Exploratory Data Analysis

After reviewing the State and Regions count plots, features of Region is found identical with State so region column is dropped.

```
df.drop('Region',axis=1,inplace=True)
```

Outliers Detection: Box plot is created

Even Though outliers are found it is not removed as it is observed that after removal of outliers ,a large amount of data is lost . (From 330 rows to 176 rows). Loss of data will not lead to good results.

Response Variable:

Adultweekday,Adultweekend,daysopenLast year,projectedDaysOpen

Identification and creation of features:

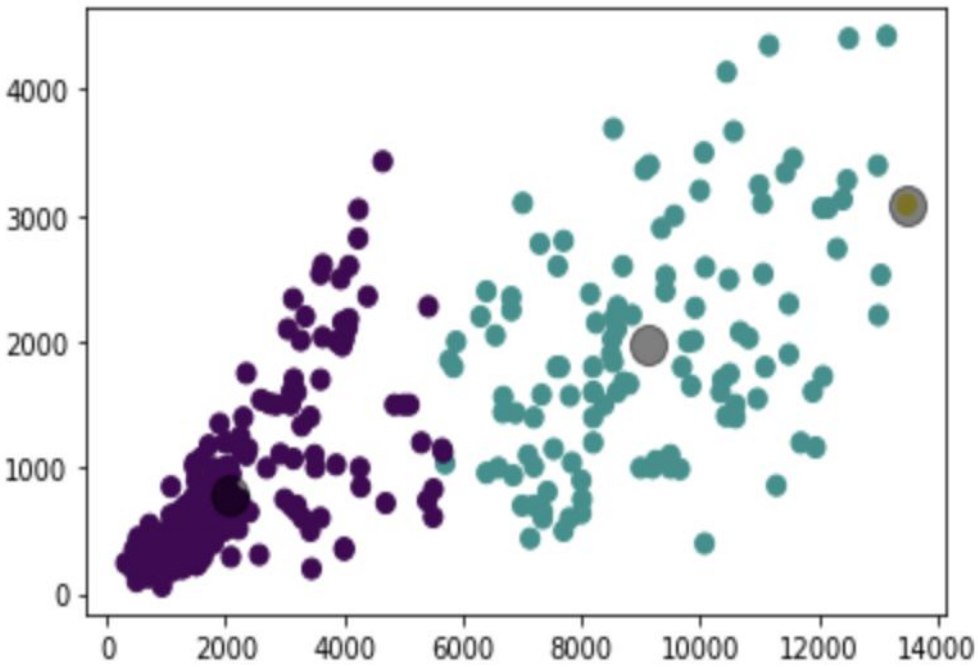
Heatmap is created to identify features that suffer from multicollinearity. The correlation matrix is displayed in a heatmap to select and remove collinear features . The feature that is more than 95% correlated is selected as base elevation and it is dropped.

Feature creation-

K-means clustering is used to develop the clusters and is used as features in model development dataset.The Elbow plot is a diagnostic tool that helps you determine the number of clusters to include in k-means clustering implementation.The error between clusters and within clusters is compared for a range of 1 to 11 clusters, and it appears the elbow is between two and four, therefore we set the parameter k = 3.

```
kmeans3 = KMeans(n_clusters=3)
y_kmeans3 = kmeans3.fit_predict(x)
plt.scatter(x[:, 0], x[:, 1], c=y_kmeans3, s=50, cmap='viridis')
```

```
centers = kmeans3.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='black', s=200, alpha=0.5);
```



Data

Preprocessing and Model Description

Dummy Variable for state is created . Using sklearn preprocessing ,scaling of the features of the dataframe is done except name of resort which is not needed in modelling ,so it is dropped.A response variable is set as Adultweekened for modeling.

```
from sklearn import preprocessing
x=df.drop(['Name','AdultWeekend'],axis=1)
y=df.AdultWeekend

scalar=preprocessing.StandardScaler().fit(x)
X_scaled=scalar.transform(x)
```

Linear Regression is used to predict price of tickets

```
lm = linear_model.LinearRegression()

model = lm.fit(X_train,y_train)
```

```
ypred= model4.predict(X_test)
```

After the iteration of models, Model4 seems to have higher accuracy and it is the best model to predict the price of tickets. Feature selection is done on the basis of feature importance and in each iteration the top features ,only actionable traits ,are taken into consideration to build the model.

Model	**Train score(R2)**	**Test score(R2)**	**Features Dropped**
Model 4	93.27	89.85	-
Model 5	91.95	87.03	state
Model 6	92.19	86.5	state,sumit_elev

Feature Importances

Big Mountain Resort should focus on actionable traits like pricing of tickets and the number of days the resort will be open. However, other traits like summit_elev , base elevation, average snowfall are not actionable. For any ski business, climate change is a big challenge as it determines the business of ski . In addition , Climate change can be done by weather forecasting , and it adds to build pricing strategy based on average snowfall as it is one of the factors which drives customers for skiing.

Review of Results

Model 4 is a generalized model as test score is more than train score and it will predict the price of lift chairs based on the model given the characteristics of the resort in

comparison to other ski resorts and their unique characteristics.The performance metrics are

Mean Absolute Error: 5.162174619564856
Root Mean Squared Error: 6.874864639122517
Mean Squared Error: 47.263763806257174

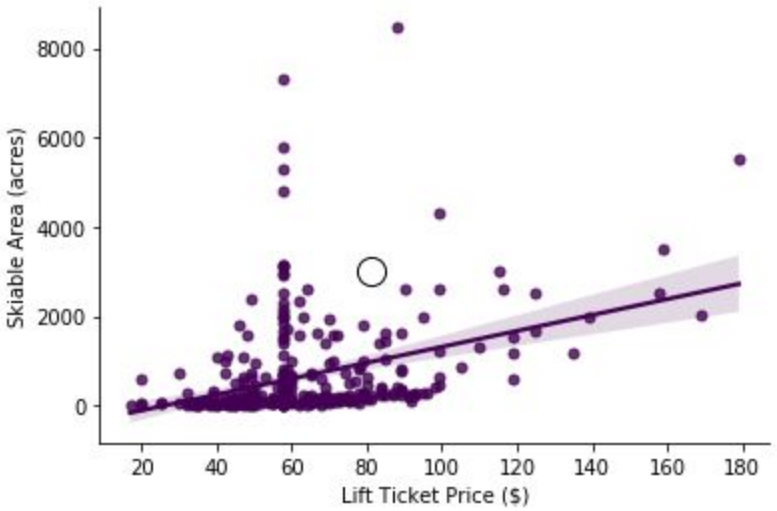
Big Mountain Resort:

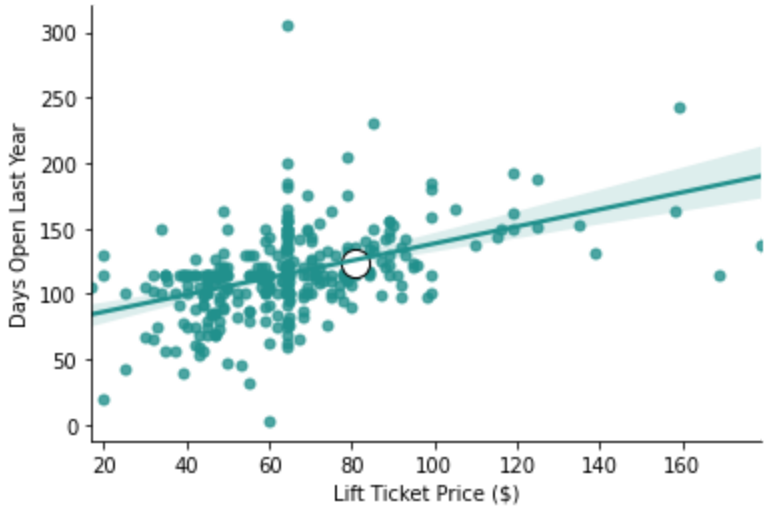
```
features=df3.drop(['Name','AdultWeekend'], axis=1)

price=  model4.predict(features)
price
```

The predicted Big Mountain Resort adult weekend price is $83.22
The actual Big Mountain Resort adult weekend price is $81.0
The accuracy of the model is 93.14% so the predicted price.

We can increase revenue considering expected price based on the model . Further, clusters can be used to create scatter plot for visualizing the Adult weekend values compared to others characteristics





CONCLUSION

Ski resorts constantly tend to conform to skier needs, achieve cost-effectiveness and generate stable revenues.

Skier needs are rated on the number of lift chairs, price, average snowfall, skier terrain and runs.

In order to meet the increased operating cost ,management needs to build a dynamic pricing strategy and increase the price of lift chairs and the number of days open in a year.

Comments:

The main purpose of the report for the data science projects is to make it interactive using the markdown language to be able to integrate it with the code.