**Problem Identification Overview**

Having increased the operational costs a ski resorts’ executives would like to figure out how to maintain their profit margin. To this end it was decided to predict the price of ski-lift tickets. A single comma separated file was provided with accompanying metadata. After inspecting and cleaning the data an unsupervised classification and subsequently a regression model were used. A presentation along with this document will serve to show the suggested ticket price and reasoning for the outcome.

**Data Preprocessing**

In preprocessing the data features showing colinearity were dropped in order to optimize the model. Outliers were observed but for the purposes of this exercise were kept in.

**Model Description**

* The data was split into train and test sets with a 75/25 split.
* Kmeans was used to create an additional feature
* Linear regression was then used as the final model for predictions
* Three model iterations were tried and with each iteration an additional feature was dropped to try to improve performance.

**Model Performance**

Table 1 shows the explained variance and mean absolute error. The closer are explained variance is to 1 the better our model. The mean absolute error shows how far our prediction was from the real values. Table 2 shows the influence of a feature on our response variable, in this case the Adult Weekend ticket prices.

Table 1

|  |  |
| --- | --- |
| Explained Variance | Mean Absolute Error |
| 0.93653 | 5.094973 |

Table 2

|  |  |
| --- | --- |
| Features | Coefficient |
| AdultWeekday | 20.12881 |
| clusters | 2.647871 |
| vertical\_drop | 2.064305 |
| triple | 1.450975 |
| Runs | 1.361873 |
| surface | 1.250484 |
| quad | 1.239127 |
| daysOpenLastYear | 1.201522 |
| averageSnowfall | 0.987277 |
| fastQuads | 0.767482 |
| total\_chairs | 0.641832 |
| projectedDaysOpen | 0.587744 |
| fastSixes | 0.465705 |
| NightSkiing\_ac | 0.423496 |
| LongestRun\_mi | 0.316646 |
| TerrainParks | 0.289076 |
| trams | 0.285837 |
| double | 0.26959 |
| yearsOpen | 0.255528 |
| SkiableTerrain\_ac | 0.124049 |

**Model Findings**

The model predicts a ticket price of ~$88 dollars based on data from several resorts.

The following figures show some of the relationships between a few features and the ski-lift ticket price. As you can see there are no obvious relationships thus machine learning techniques help us build meaningful relationships.

Figure 1.

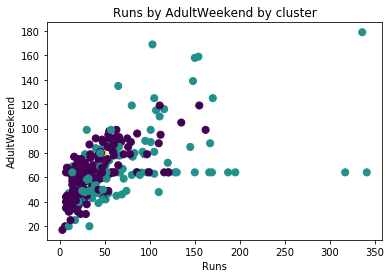


Figure 2.

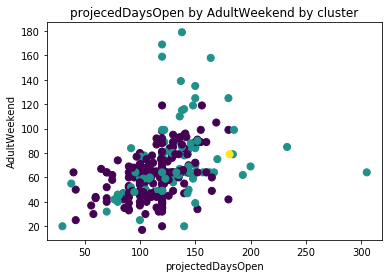
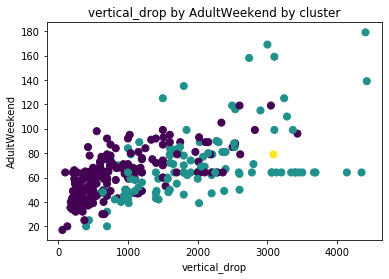


Figure 3.



**Next Steps**

The model does not answer the bigger question of maintaining a profit of 9.2%, additional data is needed. In addition, further model improvement may be needed.