Problem Statement Worksheet (Hypothesis Formation)

What opportunities exist for Big Mountain Resort (BMR) to recoup this year the operational costs ($1,540.000) from the chair installed and maintain a profit margin of at least 9.2%? What is the estimation of the year's annual revenue be after these opportunities are exploited?

1 Context

BMR, opened in 1947, is a Resort is located in NW Montana, it offers spectacular views of Glacier National Park and Flathead National Forest. BMR has an annual inflow of 350.000 skiers and has recently installed a new chair that has increased operational costs in $ 1.540.000.
BMR has an annual profit margin of 9.2% and is interested in **identifying opportunities to recoup the operational costs of the new chair** to maintain this margin. Additionally, BMR is interested in understanding **what the annual revenue will be** after exploited the opportunities identified.

2 Criteria for success

To (i) **identify and implement opportunities** that allow BMR to **maintain a profit margin of 9,2%,** this is, to generate through cost reduction, new revenue or a combination of both, a minimum of $1.540.000, and (ii) **estimate the annual revenue** after implementing these opportunities.

3 Scope of solution space

The focus is to **find and implement ways to increase revenue, reduce costs, or a combination of both, to maintain a profit margin of 9,2%.** To do this, we will analyse the data of potentially 330 ski resorts in the USA to find insights regarding relationships between the price of these ski resorts and other factors. These insights will allow us to understand what factors are more determining to the price, which could help business to find opportunities in reducing cost or increasing revenue.

4 Constraints within solution space

- The **estimations will be based on historical data**, and therefore, deviations of real data vs. projected data could affect the projections.
- It is necessary to ensure that there will be **access to the Database.**
- The opportunities need to be **identified and implemented in the current year**.
- The opportunities need to be **measurable,** in order to estimate the annual revenue after implementing them.
- Some ski resorts may have missing or wrong data. Therefore, we will need to either delete or impute the missing fields, which could impact negatively our assumptions and further conclusions.

5 Stakeholders to provide key insight

Jimmy Balckburn – Director of Operations
Alesha Eisen – Database Manager

6 Key data sources

CSV file – Metadata with information about 330 resorts in the US that can be considered as BMR competitors, the data provides information about different factors of ski resorts, for example:

i) **Structural data** (annual inches of snow, nº of skiable acres, ..)
ii) **Utilities data** (nº of chairs of different kinds, nº of ski tracks, ..)
iii) **Operations data** (prices of chairlift tickets, nº of days opened last year, number of days projected to open next year, …)

For our analysis, we will focus on price ticket as the variable to study based on the rest of the available variables in ski resorts.

Recommendations and Key Findings

- **The model supports that BMR may be under pricing**
- **Only 8 factors seem to be the ones that better explain the ticket prices of ski resorts**
- **None of the opportunities analysed can by themselves recover the cost of the chairlift**
- **Adding the expected operational cost increase/reduction of each opportunity is necessary to compare and select the best opportunities to implement**

1 BMR may be under-pricing

According to the information provided, BMR is charges $81 for the ticket price.

However, the model expects that the price of BMR ticket is between $81.09 and $101.73, but most probably around $91.41.

Therefore, it could be reasonable to state that BMR may be under pricing. However, we are not awake of particular circumstances of the ski industry that may be determining a lower price. On the opposite, our assumptions are based on that the ski market is operates as a free market.

2 Factors that better explain ticket prices

In the data provided there were several variables included for each ski resort. The model that best predicted the ticket price considered that 8 factors alone best determined the ticket price.

The factors that the model considered that best explained the ticket prices are:

- Vertical drop
- Area covered by snow making
- Nº of chairs
- Nº of fast quads
- Nº of runs
- Longest run length
- Nº of trams
- Total skiable area

3 Scenarios / Opportunities analysed

A total of 4 scenarios were tested to observe how the ticket price was expected to change based on additions or decreases in one or more of the 8 variables defined.

- **Close up the 10 least used runs:**
  - close 1-2 runs → expect -$,88 in ticket price
  - close 3-5 runs → expect -$1.3 in ticket price
  - close 6 runs → expect -$1.5 in ticket price
  - close 7-10 runs → expect -$2.1 in ticket price
- **Increase vertical drop in 150 feet + install a chairlift:**
  - expect +$1.24 in price increase → + revenue = $2.17M + ops cost = $ 1.54M,, net revenue = $ 0,63M
- **Same as previous + add 2 acres of snow making:** → no further expected ticket price increase
- **Increase longest run in 0.2 miles + add 2 acres of snow making:** → no expected increase in ticket price at all

4 Improvements to the model (ops costs)

The model does not include information about the operational savings from each opportunity. Therefore, the analysis is not complete. In particular, it only gives information about the expected price increase or decrease, but not the operational cost increase or savings from this option.

To better compare alternatives it is necessary that business integrates the information of operational costs to the analysis.

Depuration of data

The data provided has many missing observations as well as different missing observations for certain variables. In order to prepare the data for modelling, we first depurated the data, by either deleting ski resorts with missing data or imputing certain factors (with the median and mean)

1 Initial situation – Raw Data

The original data contained 330 rows, each one of them represented a different ski resort in the United States.

2 Modifications performed to the Data

We decided to delete the column that contained fast eights chairlifts for all the set due to having half of the values missing, and the rest of the values with zero value. Only one resort had a value of 1.

We decided to check online the value of one resort's skiable terrain area as it was suspiciously high. We observed that it was a missing value a replaced it with the value we web scraped from their website. It seemed to be a translation error, as part of the number matched with the web data.

One resort had a value for years open of 2019, we were not sure if this meant that it had been open in 2019 and therefore it should have a value of 1. As minimum value for this factor in set was 6, we decided to delete the row.

Our target value was ticket price, there were two fields that showed information related to ticket price. Ticket price week and ticket price weekend. We counted how many rows had one missing or both missing. We deleted all rows with the 2 values missing (14% of set). With the rows left, we checked how was the distribution of week price and weekend price for all resorts. Resorts with prices above $100 had same prices for week and weekend, also in Montana prices were the same for week and weekend tickets.

Therefore, we took to the decision of including weekend prices as the variable to study. As a result of this, we deleted the columns of week prices in its entirety.

As a result of al the amendments performed, our dataset ended up with 277 rows instead of the original 330 rows.

Exploratory Data Analysis

To gain a better understanding of the industry situation, we decided to generate and additional data set based on the factors state and Region from the original data set. In particular, we downloaded information about the population, area and other factors for each combination of State and Region, to further analyse their particularities.

1 Source of the data

We gathered the data from the following source:
https://simple.wikipedia.org/wiki/List_of_U.S._states

2 Content gathered from the data

By adding data about the area and the population of each state, we were able to generate an number of interesting state-wise statistics of the ski industry in the United States.

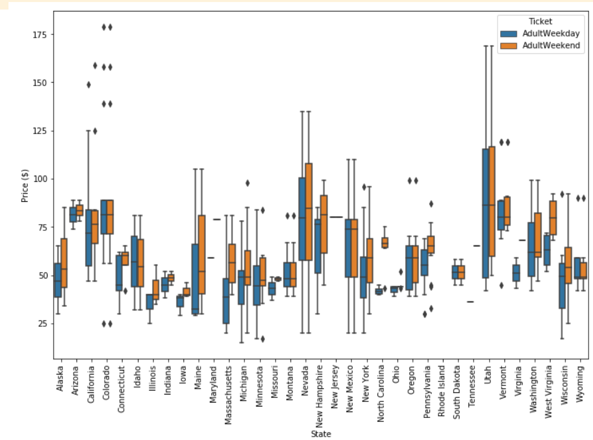
For example, we analysed the following variables:

- Total state area
- Total state population
- Nº of resorts per state
- Total skiable area
- Total night skiing area
- Total days open

3 Metrics calculated from data

The generation of this additional dataset allowed us to calculate certain statistics that were useful to better understand the industry of ski resorts, such as:

- Resort density
- Average ticket price of states (top right)
- Feature correlations (bottom right)



From Baseline Model to Final Model

We firstly defined a baseline model using the average price of the dataset as predictor of price, which led to an average error of $19.
Secondly, we tried different models and selected the one with best accuracy (Random Forest Regressor), with an average error of $9.66 and std of $1.35.

1 Baseline Model – Simple Average

To set a baseline against which to compare our models performance, we first tried to estimate the ski resorts ticket price by using the simple average of the dataset.

The accuracy of the model was such that you could expect to be off the actual price, on average, in $19.

2 Imputing values / Scaling values

As the dataset still contained many missing values for rows, we had to make a decisions of how to impute the missing values. We considered the approach of using the median of the dataset as well as the mean of the data set to check which one provided with better results.

Also, we checked what the accuracy of the models was scaling the variables and without scaling them.

3 Linear Regression Model

The first model we tried as a Linear Regression model.

The model showed an accuracy of $10,50 with a std of $1.62 in the training set.

The model showed an accuracy of $11.79 on the test set, which is consistent with the previous results

4 Random Forest Regressor

The second model we tried was a Random Forest Regressor.

The model showed an accuracy of $9.66 with an std of $1.35 in the training set.

The model showed an accuracy of $9.50 in the test set, which is consistent with the previous results.

5 Conclusions

As the **Random Forest Regressor** demonstrated a better accuracy than the Linear Regressor, we decided to choose it as the model to keep working on with.

A small remark is that Random Forest Regressor did not consider *Longest run length* as a useful variable to explain the ticket prices of ski resorts, while Linear Regression model did.

However, considering that the accuracy of the Random Forest Regressor could estimate the price up to $1 better on average than the Linear Regression model, we considered that it was worth to keep the second model with its assumptions as the strongest.

Summary and Conclusion

The opportunities analysed are not complete unless the information about operational costs is integrated in the analyses and conclusions. The model determined that there is room to increase the ticket price of BMR and that none of the opportunities analysed can recover the cost by themselves. Further investigation from business is needed.

1 Capabilities of the model

The model generated can predict the expected increase or decrease in ticket price for BMR or other ski resorts with an accuracy of +/- $9.66 and std of $1.35.

In practice, business can input scenarios on the 8 most relevant features of the ski resort, this is, adding or discounting nº of chairlifts, nº of runs, area of snow making, … and expect in return the expected increase or decrease in ticket price from these modifications.

2 Limitations of the model

The model is not capable to include in its predictions the expected increase or decrease in operational costs form the scenarios that are input into it.

As a result, if a business stakeholder wants to know the total impact in terms of revenue of a particular scenario (e.g. closing 5 runs and adding 1 chairlift), the model will predict a decrease in expected ticket price from closing the runs, and an increase in ticket price from adding the chairlift.

However, the business stakeholder will need to analyse in parallel what will be the operational cost savings from closing such 5 runs, and what will be the operational cost increase from installing the new chairlift.

Following this approach, business will have the total picture of the scenarios and will be able to plan the best strategies to achieve the results planned.

3 Scenarios already analysed

As commented in previous slide nº 2 of this presentation, none of the scenarios input in the model by themselves can recover the cost of the chairlift investment performed.

However, as the model does apparently support enough a ticket price increase of around $10 on average on price tickets, it would be interesting to understand if BMR may be under pricing their tickets, and to what extent it is a possibility to increase the ticket price without losing clients, or losing as few as possible so that the price increase is worth it.

At the same time, due to currently being unaware of the impact on costs of the scenarios, the first thing to do would be to include this information in the analyses and see if the reduction of operational costs can compensate the loss from expected ticket price resulting from some scenarios.

4 Next steps

Business stakeholders should input to the model different scenarios, modifying the most relevant variables, aiming at finding the best opportunities to optimize revenue.

As expected Price Will be given by the model, their task is to add the information of expected operational costs to gain the full economic understanding of each alternative.