In advance of your submission, please:

1. Complete the following details and include as the first page of your submission;
2. Read the Turnitin section and check that you understand how Turnitin is used to assess your work;
3. Read the declaration to check that your submission conforms with the listed requirements before you submit your work.

|  |  |
| --- | --- |
| Name(s) and Student Number(s): | Deeksha Deepak |
| Module Title: | Predictive and prescriptive analysis |
| Title of Coursework: | Management and Business analytics |
| Word Count  (see declaration below): | 2855 |

|  |
| --- |
| **Required Format**  Font: Verdana  Points: 10  Paragraph line spacing: 1.5  Page Numbers: Included |

|  |
| --- |
| **Turnitin Similarity Check**  Where appropriate to the coursework assignment, your document will be submitted to Turnitin to generate a similarity report for review by your tutor. This report will compare your work against millions of previously submitted student papers and online resources (e.g. web sites, journal articles) in the Turnitin database and highlight any text that matches your submission. |

|  |
| --- |
| **Declaration**  By submitting your work you are certifying that:   1. The submission is the result of your work and does not contravene the University Regulation on Academic Irregularities. 2. The word count included on this cover sheet is accurate and follows the guidelines outlined in the assignment brief (failure to include an accurate word count will be treated as a minor academic irregularity as defined in the Academic Irregularities Policy.) 3. Your ability to complete your assessment has NOT been adversely impacted by circumstances beyond your control. 4. Once you have submitted your work, any such circumstances would need to be disclosed through the Academic Appeals Policy and process and not through the Notification of Extenuating Circumstances Policy and process. |

Table of Contents

[2 INTRODUCTION 5](#_Toc99244457)

[3 PREDICTIVE ANALYSIS 6](#_Toc99244458)

[3.1 REGRESSION MODEL 7](#_Toc99244459)

[3.2 ANOVA TEST 10](#_Toc99244460)

[3.3 CLASSIFICATION TREE 11](#_Toc99244461)

[4 PRESCRIPTIVE ANALYSIS 12](#_Toc99244462)

[5 CONCLUSION 15](#_Toc99244463)

[6 RECOMMENDATIONS 16](#_Toc99244464)

[7 BIBLIOGRAPHY: 17](#_Toc99244465)

[8 APPENDICES: 18](#_Toc99244466)

**EXECUTIVE SUMMARY**

The report focuses on helping Kate and Joe owners of the hotel Chrysalis located in Las Vegas in deciding whether the hotel only or the hotel and casino need to be renovated.

The report’s goal is to increase the occupancy rate of the hotel by renovating the facilities available. Using the reviewer’s score from TripAdvisor is indirectly dependent on the facilities available in different hotels around the world accordingly, to recommend which facility works best with the hotel to increase its profit and footfall by using predictive and prescriptive analytical techniques. A regression analysis was performed on multiple model alternatives which showed that the two best facilities that should be considered are Free internet and pool with a p-value of less than 0.05. After studying the regression model and utilizing the Anova to find the best model and evaluation to emit possible errors 1.058 and 0.859 (RMSE and MAE) as shown in figure 8 it is seen that the best decision is to consider renovating the hotel, with THE pool and free internet. It is recommended to renovate the gym and replace the slot machines as well as it would offer better chances to increase the occupancy rate. A classification tree was also conducted to see how many people from the dataset prefer having either the pool or free internet or both and how it would directly affect a hotel’s score. For prescriptive analytics, Decision theory was conducted to find the profit earned by the hotel after deducting its expenses on the renovation from the revenue. After analysing the calculations, it is recommended that the hotel only is renovated and slot machines are replaced as the profit ($14,560,000) earned is more compared to renovating the hotel and casino also considering that the EVPI= 0. It is also recommended that Kate and Joe consider using the whole dataset to take a decision as customer satisfaction is the primary key focus for any business. By choosing to renovate the hotel and replacing the slot machines the occupancy rate of the hotel is very likely to increase to 80% and bring more footfall for the casino as well. This next step of renovation will help the hotel that was built in 1965 be revamped and fully equipped for the present age attracting a lot more customers from around the world to gamble and stay in Las Vegas although the hotel is located downtown and not on the strip.

# INTRODUCTION

The following report has been conducted on the analysis of Hotel Chrysalis using 2017 data reviewers’ opinion of Las Vegas hotels found on trip advisor. The report contains analysing, cleaning, modelling and statistical techniques using the trip advisor dataset to come up with the perfect decision. This report aims to use prescriptive techniques to determine whether the owners Joe and Kate should consider renovating only the hotel or both the hotel and the casino fully. Regression model, Anova analysis has been used in regards to predictive techniques to answer the question that is whether Joe and Kate should renovate all the facilities or just a few. Analyses exploring the difference between the facilities considered to renovate in comparison to the revenue earned by the owners are also included. An assumption has been made that the revenue earned by the hotel and casino is enough to cover the facilities and renovation costs after their usual expenses.

**DATA STRUCTURE AND UNDERSTANDING**

The dataset shows different facilities like a gym, pool, spa and free internet provided by other hotels and casinos in Las Vegas and how the customers feel about the facilities being available to them during the duration of their visit. As most of the relevant data in the dataset are in a categorical format, the independent variables have been changed to the quantitative form of 0s and 1s. The facilities columns have been renamed and mentioned below Pool.1, Gym.1, Tennis\_court.1Spa.1, Casino.1, Free\_internet.1. There is some data missing under the column names user continent, member years, review month, review weekday which has not been altered or removed because it has not been used for any of the required calculations.

A researcher can use descriptive statistics to quantify and characterise the basic properties of a data set. As a result, descriptive statistics serve as a foundation for data analysis, allowing researchers to organise, simplify, and summarise information (Allen, 2017)

There is a potential that the data obtained is not always pure and contains impurities. The data must be thoroughly cleansed before analysis can begin. Depending on the type of dataset issues and insights, data cleansing usually includes a variety of solutions. Each strategy has advantages and disadvantages. Incorrect data is either fixed, modified, or eliminated, making analysis more straightforward.

**Research Questions:**

1. What facilities should be considered to increase customer footfall?
2. Which renovation option should Joe and Kate decide on implementing, the hotel and casino or hotel only?
3. Should Kate and Joe’s decision is based on all reviews or just a sample of them?

# PREDICTIVE ANALYSIS

Predictive analysis' significance stems from its capacity to advise the optimal future planning by integrating data on who, what, where, and when to investigate why and how. Predictive analytics is the use of historical data, machine learning, and artificial intelligence in the workplace to forecast future events. (Henrys, 2021)

‌This part of the report focuses on analysing the data and deciding which facility should be considered for the renovation of the hotel and/or both hotel and casino. Machine learning software like R studio was used to predict the results explained below. To help the owners Kate and joe make a better decision, models like the multiple linear regression model, ANOVA model for validation, decision tree have been used.

## REGRESSION MODEL

Predictive analytics model Regressions has been used to predict the best decision for Kate and Joe to take as their next steps for their hotel’s renovation. Various tests on the correlation between Score taken as the dependent variable (Y)and the facilities like pool, spa, gym, free internet, tennis court, and casino taken as independent variables (X) was conducted to find the most suitable outcome for Hotel Chrysalis. The tested models have been explained below

Model 1 :

In this model, the dependent variable(Y) is Score and the independent variables(X) are Pool, Gym, Tennis court, Spa, Casino, Free internet. It is observed that there are two significant values, i.e “Pool.1” and “internet.1” having significant p-values less than 0.05. It is also seen in Figure 1 that Gym, Tennis court, spa and casino have no significance showing its P values are more than 0.05 and the adjusted R-squared when the codes run is 0.07503. On the other hand, more models were run where facilities that have p-values more than 0.05 were removed individually for better decision-making.

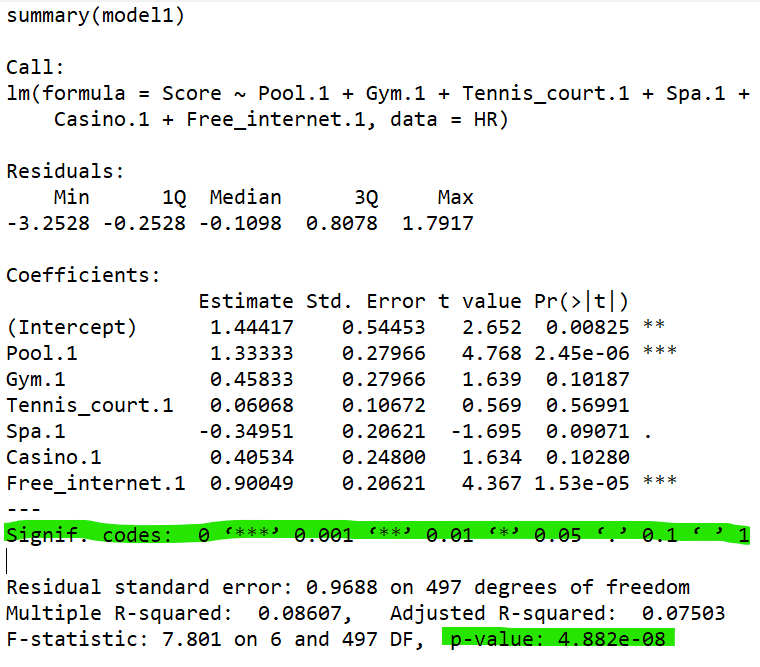
****

Figure 1

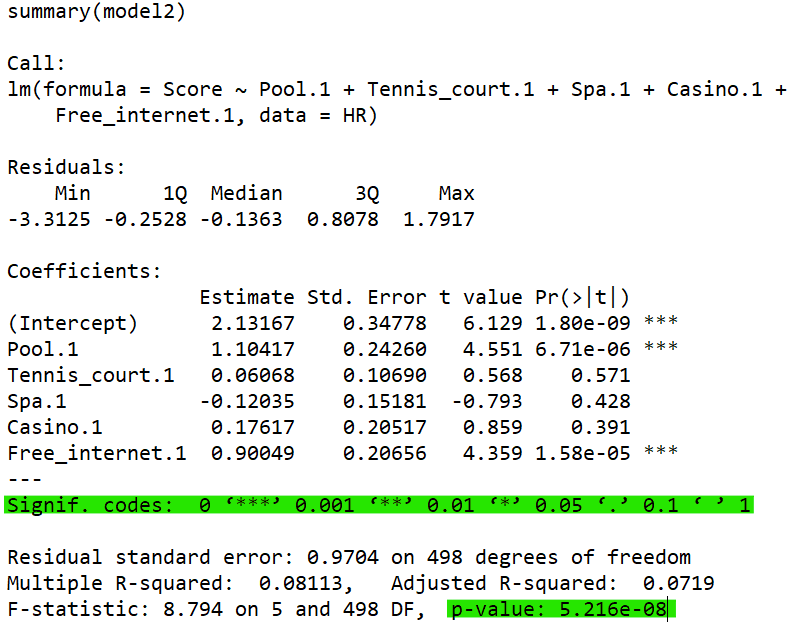
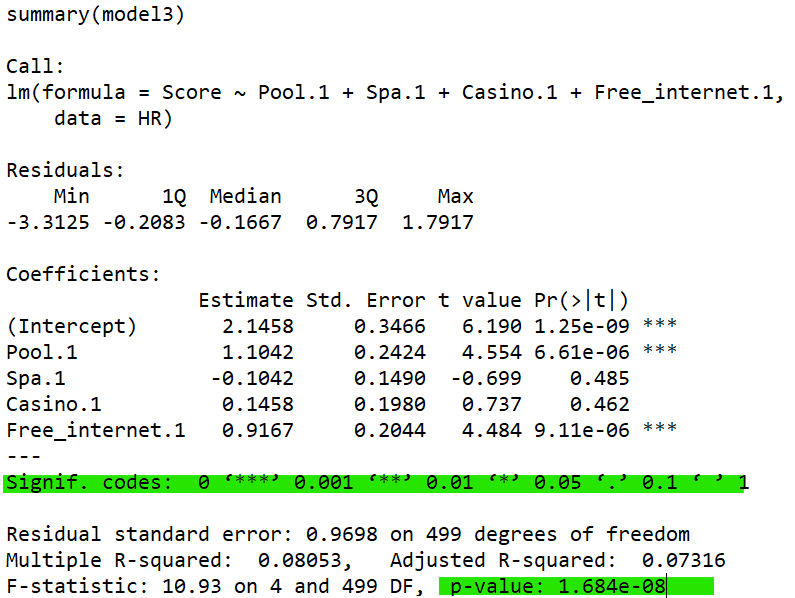
****

Figure 2

**Model 2:**

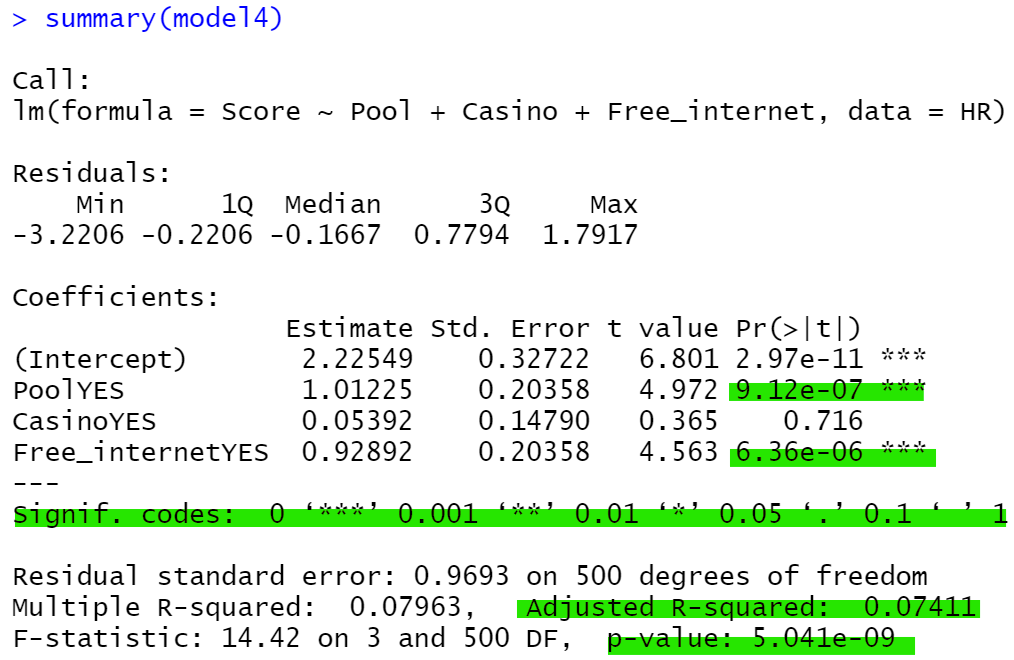
To have a better understanding of the regression model, one of the unimportant independent variables was removed while keeping the dependent variable constant. In model 2, the variable Gym was removed, and the model was then re-run, yielding an R-squared value of 0.08113, indicating that there was little change. The P-value stays below 0.05, showing that the pool and free internet remain superior estimators.



**Model 3:**

Another model was created by deleting a non-significant variable, in this case, the Tennis court, yielding an R-Squared value of 0.07316, which is somewhat better than the prior model. When translated from scientific form, the p-value of pool and internet is still less than 0.05 when compared to other variables, demonstrating that they are still the best estimators.

Figure 3



**Model 4:**

The variable “Spa” was eliminated for this model to get an estimation of the r-squared value to be 0.07411 showing a rise. The P-value for the facilities pool and free internet remain less than 0.05 proving they have a better score and also rejecting it as a null hypothesis because it proves that the variable X is significantly related to Y. It also shows that it has good significance with the asterisks.

Figure 4

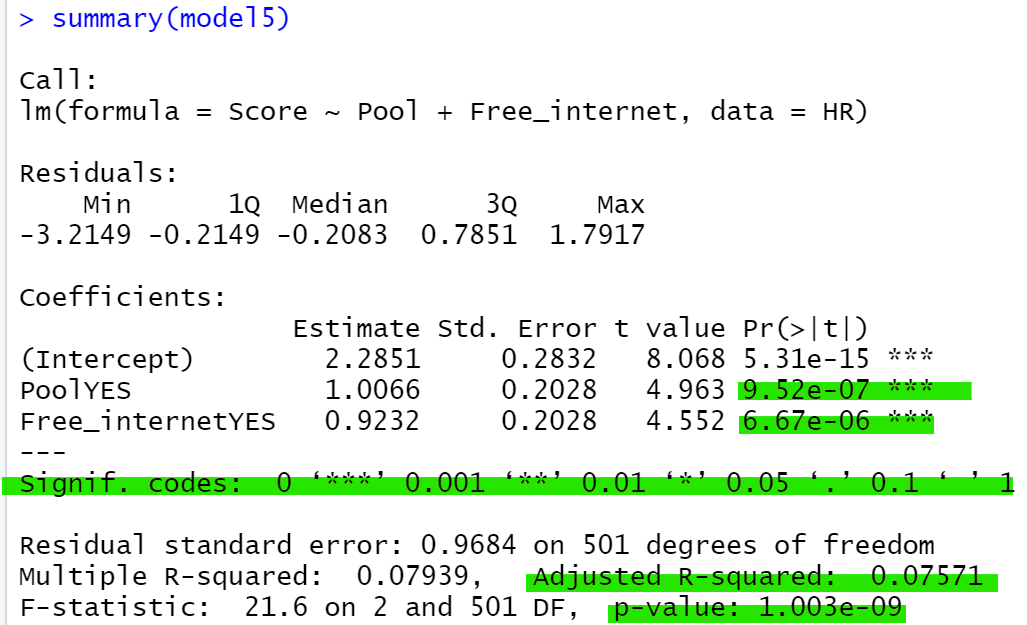
**Model 5:**

Figure 5

The final model for regression by removing the last insignificant variable “Casino”. In this model with pool and free internet as predictors, the adjusted R-squared is 0.07571. This means that 7.57% of the variation in score can be explained by the pool and free internet. After removing casinos from the model, the adjusted R-square increases to 0.07571.

## ANOVA TEST

To choose which model works best multiple ANOVA tests were conducted between models. ANOVA is used to compare different group means in a data set. ANOVA was run on model 1 and model 2 up until Model 5 to choose which fits best based on p-value and then conclude whether it is a complex model or a simple model.

Once all the models were compared, it has been concluded that model 5 shows a significantly better fit to the hotel compared to model 1 showing a p-value of more than 0.05 which is 0.7156 as shown in figure 6 proving that it is a simpler model and it is more favourable. Therefore, we should reject model 1 and choose model 2 with pool and free internet as its independent variables.

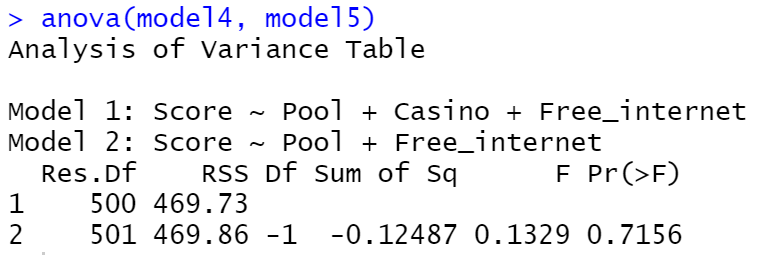
****

Figure 6: ANOVA MODEL

To compare and evaluate the prediction errors of the model’s actual value and the predicted value, root mean square error and mean absolute error were calculated. The root mean square error, a statistic that specifies the average gap between the model's predicted values and the actual values in the dataset, is important to analyse how well a regression model fits a dataset. The smaller the RMSE, the better a model fits the data.

The packages ‘CARET’ and ‘ISLR’ were first installed before running the prediction RMSE and MAE code. The single train set split was also run stating that 80% of the data goes in training and the remaining 20% goes in validation, as shown in figure 7. Considering that Pool and free internet have the best significance codes, an RMSE code and MAE code was run resulting in RMSE being 1.058 and MAE being 0.859. The **RMSE** value tells us that the average deviation between the predicted score made by the model and the actual score is 1.058 which is considered as high and the r-squared value is 0.07571. Since the MAE code resulted as 0.859, it is quite low, therefore predicting a better model fit.

Although the RMSE and MAE were calculated the data given is insufficient as there are no particular variables that help in the exact prediction. Variables like housekeeping, food served at the hotel, customer service and so on are huge factors that directly affect the score of a hotel. Therefore making the prediction inaccurate.

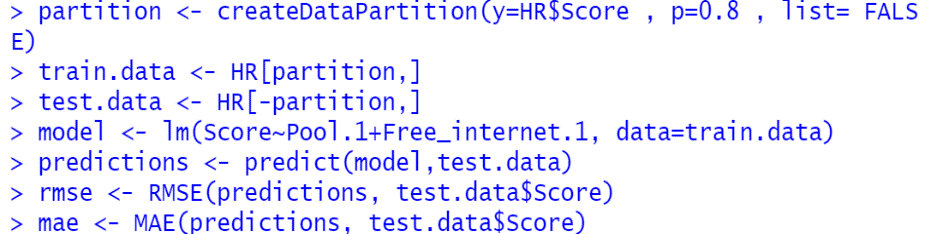


Figure 7

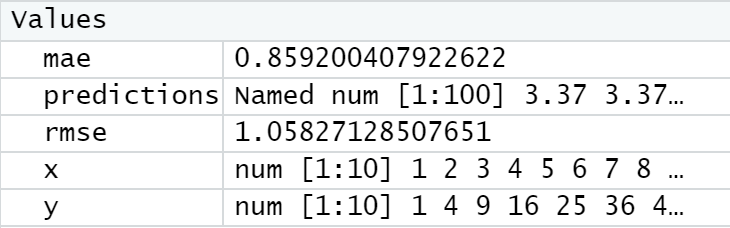
****

Figure 8

## CLASSIFICATION TREE

A classification tree was built using the independent variables that were used to conduct the regression models earlier in the report. The classification tree states which is the best combination comparison to the factors between individual variables.

The tree predicts whether the score of the hotel is affected if the pool is available at the hotel or there is free internet provided to all the customers or both.

The average score of the hotels in the parent node is 4.1, with a total of 504 instances representing customer evaluations on Trip Advisor. The left child node reveals that if there is no pool, the hotel's score drops to 3.2, while 24 cases (or 5% of the data set) are OK without one. The score climbs to 4.2 in the right child node, with 480 occurrences or 95% of the sample requiring a pool for their stay at the specific hotel. If both pool and free internet are included the prediction changes, the tree classifies out of the 480 people who want a pool, 24 people or 5% of the dataset don’t need free internet, but the other 456 customers prefer having a pool and free internet which in conclusion states that majority of the customers who have reviewed on trip advisor consider a hotel in Las Vegas that provides free internet and also has pool access. If a facility was not available, the hotel's score would be dramatically affected; for example, if there was no free internet, the score would drop from 4.2 to 3.3; however, if it had free internet plus a pool, the score would stay the same or even grow. It is also seen in table 4 that the leading reasons to visit Las Vegas is for vacation/pleasure therefore it is best to have the amenities refurbished and renovated for best customer satisfaction and to upkeep and increase the score of the hotel.

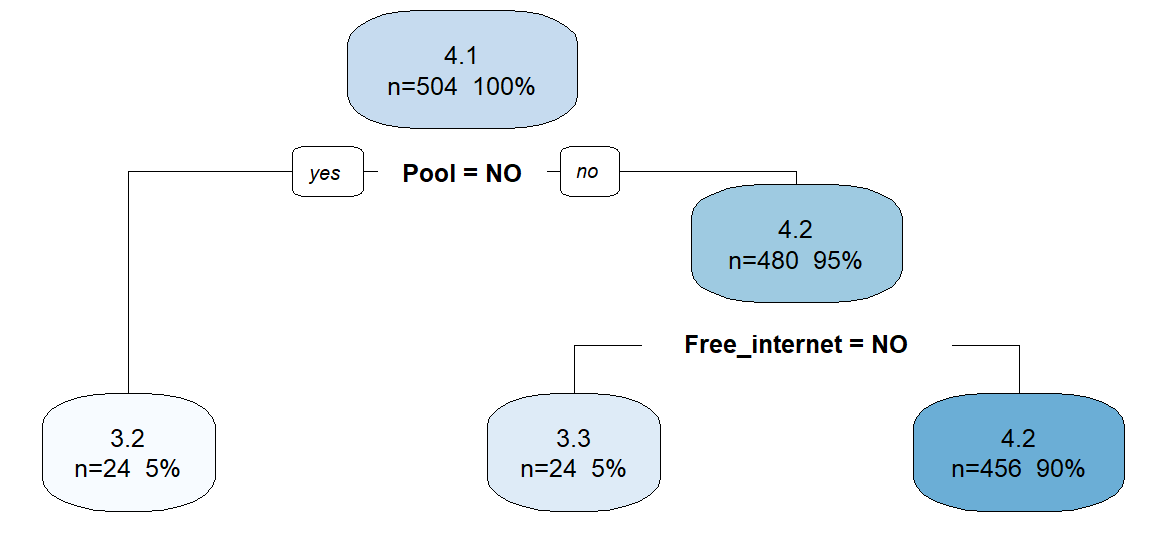


Figure 9

# PRESCRIPTIVE ANALYSIS

While predictive analysis allows a company to make future decisions using past data, prescriptive analysis is used to make the best suitable decision using available constraints. Therefore, to help Kate and Joe take a decision whether they should renovate the hotel or both hotel and the casino a decision theory was conducted.

## Decision Theory

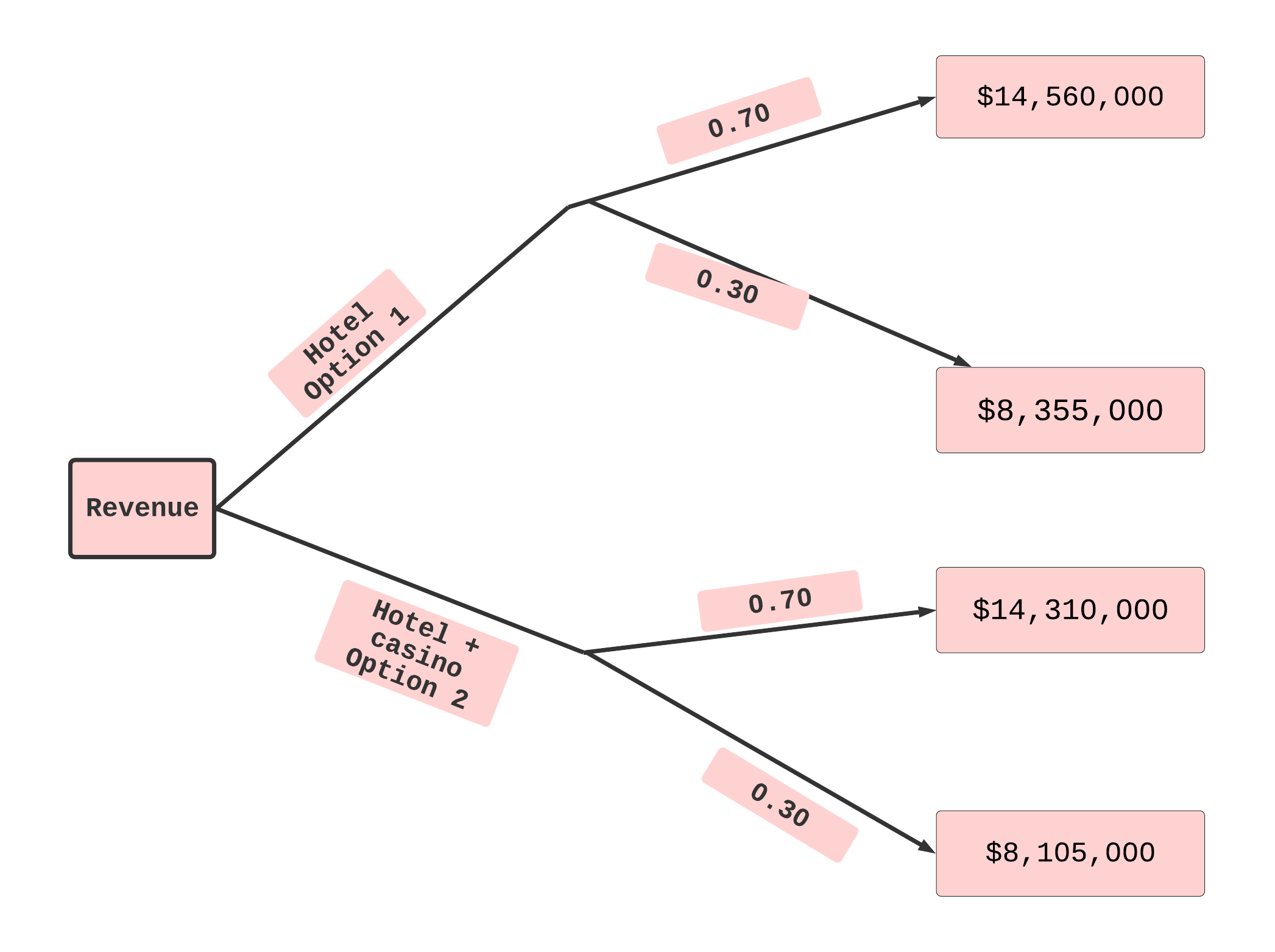
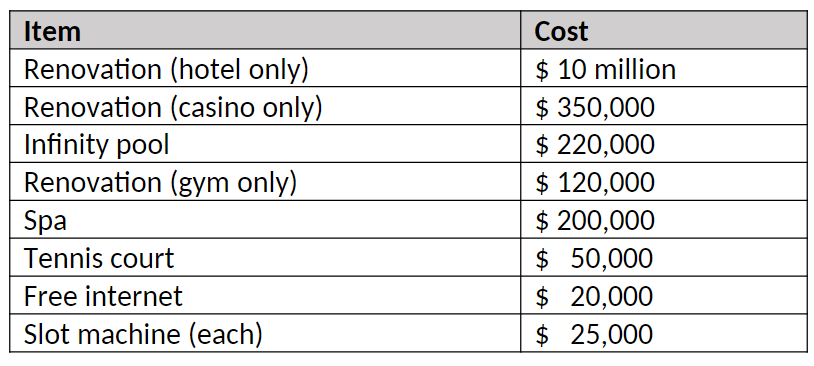


Figure 10: Decision Theory

According to the case brief, the hotel has 1000 rooms and each room generates a profit of $60. It also states that if Kate and Joe decide to rebuild their hotel, there is a 70% probability that footfall would grow to 80% and a 30% chance that it will remain unchanged. With the 70% and 30% in mind, more calculations have been done to aid in the decision-making process if Joe and Kate must select option 1 or option 2. There's also the issue of the hotel's facilities and whether to remodel the slot machines and the casino or just the slot machines and a few other facilities.

Table 1



No. of rooms : 1000

Calculations:

**OPTION 1: Hotel**

* **70%**

Yearly Revenue of Hotel : 1000\*80% \*% 85\* 365 = $24,820,000

Casino Revenue: $600,000

Total Revenue (Hotel + Casino): $25,420,000.

* **30%**

Hotel revenue: 1000\*60%\*$85\*365 = $18,615,000

Casino Revenue: $6,00,000.

Total Revenue (Hotel + Casino):: $19,215,000.

Total Expenses: Hotel+ Pool + Free internet + Gym + Slot machines.

$10,000,000+ $220,000 + $20,000 + $120,000 + $500,000

= $10,860,000

Predicted Profit Earned:

Profit = Revenue – Expenses.

Profit 70% : $25,420,000 - $10,860,000

= $14,560,000.

Profit 30%: $19,215,000 - $10,860,000

=$8,355,000.

**OPTION 2: Hotel + Casino.**

* **70%**

Yearly Revenue of Hotel : 1000\*80% \*% 85\* 365 = $24,820,000

Casino Revenue: $600,000

Total Revenue (Hotel + Casino) : $25,420,000.

* **30%**

Hotel revenue: 1000\*60%\*$85\*365 = $18,615,000

Casino Revenue: $6,00,000.

Total Revenue (Hotel + Casino): $19,215,000.

Total Expenses: Hotel+ Pool + Free internet + Gym + Slot machines + Casino.

$10,000,000+ $220,000 + $20,000 + $120,000 + $500,000 + $35,000.

= $ 11,110,000.

Predicted Profit earned:

Profit = Revenue – Expenses

Profit 70% = $25,420,000 - $ 11,110,000

= $14,310,000.

Profit 30% = $19,215,000 - $ 11,110,000

= $8,105,000.

EV(Hotel) = 0.7($14,560,000.) + 0.3($8,355,000)

= $10,192,000 + $2,506,500

= $12,698,500

EV(Hotel + Casino) = 0.7($14,310,000) + 0.3($8,105,000)

= $10,017,000 + $2,431,500

= $12,448,500

**->EVPI**

EVPI=Expected payoff under certainty-Expected payoff under risk.

Expected payoff(Revenue) under certainty – Expected Payoff under risk

= [0.7($14,560,000)+ 0.3($8,355,000)] - $12,698,500

= 0

According to the calculations it is seen that the profit earned by renovating only the hotel including the slot machine is more than the profit earned by renovating the hotel and the casino. As EVPI= 0 the decision the owner’s Kate and Joe take will remain the same with no change in the outcome of EVPI. The cost of renovation for the hotel, the pool, free internet and the gym has been included also the slot machines considering it is a part of the hotel and not the casino renovation has been considered. The cost of the pool and free internet was considered because the regression model as discussed above showed a good level of significance compared to other facilities. The hotel has two gyms and renovating one of the gyms has been considered because the data in the dataset when compared to other facilities excluding pool and free internet shows that out of 504 reviewers, 480 members have said yes to the gym proving the majority of the people need a gym at the hotel. Also, the replacement of 20 slot machines has been considered because Las Vegas’s main footfall is for gambling and slot machines play a vital role in it. This would benefit the hotel’s customers reach as well.

# CONCLUSION

A data collection from trip advisor reviewers on Las Vegas hotels was examined to arrive at an analytical judgement. The data collection includes details on which hotels offered certain amenities, such as a pool, gym, free internet, spa, and tennis court. With these facilities, a study was undertaken to evaluate if the hotels' overall score would be influenced by the lack of them.

Multiple regression models were run to find the best model with only significant variables. Specifically, in this dataset being Pool and Free Internet with a p-value less than 0.05. To find more clarity with the results of the regression model an evaluation was performed using a single training/test split was run to show the RMSE and MAE values low accordingly (Figure 8) when compared to other models. Although most of the information given in the dataset focuses on the facilities provided directly affecting the score, a lot of other factors also need to be considered like customer service etc, to upkeep the score of the hotel and make improvements where necessary. The ANOVA model was also conducted to see the best fit out of the models to help influence in taking the best decision by using the R-squared value.

To get more insight on who would rather prefer only free internet, only pool or both a classification tree was formed to see how the score would vary with the lack of facilities on one hand or if the facilities were provided.

To include prescriptive analysis to the data for better decision making, a decision theory was calculated by comparing 2 options, the hotel only or both hotel and casino. The revenue was calculated with the possibility of 70% chances that the occupancy rate will increase to 80% or 30% chance that it will remain stable (60%). The payoff yearly revenue calculated for the hotel+casino is more. EVPI was also calculated to be 0. The costs included to renovate the hotel as per its immediate need was the infinity pool, Free internet, Gym and slot machines. Although built-in 1965 the case brief states that the casino does not need any immediate renovation it has not been considered. Therefore, in conclusion considering the calculated profit for the hotel is more than for hotel and casino, the most reasonable decision for Kate and Joe would be to choose to renovate only the hotel and the slot machines.

# RECOMMENDATIONS

1. Looking at the regression models and ANOVA the output received shows that the pool and free internet have good significance (7.57%) to increase the occupancy rate of Hotel Chrysalis. Therefore, it is recommended that Kate and Joe consider renovating these two facilities. The dataset also states that most people prefer a gym in the Hotel during their visit so they should consider renovating one of the gyms for now and another one later. The classification tree also shows that the score is affected by the availability of free internet and a pool both at the hotel.
2. From the decision theory, it is seen that renovating the hotel is a necessity as it is not in the best shape. It also shows that the profit earned by renovating the hotel only is more than renovating the hotel and the casino directly increasing the occupancy rate of the hotel. It is best recommended to focus on renovating the hotel including the gym, infinity pool and free internet. 73% of people remained on the Vegas Strip, according to Las Vegas gaming revenue statistics from 2017. The same year, however, 40% of the total players visited downtown Las Vegas to gamble. This figure indicated a noteworthy increase of 31% compared to 2015. (Anon, 2021). This shows that replacing the slot machines would be prime for the hotel to get more customers to stay and use their casino.

1. According to the bar graph (table 2), the majority of continents provide pool and internet to their customers in order to provide total satisfaction. As all of these evaluations and calculations are predicted utilising all of the reviews, Joe and Kate should consider the entire dataset. For improved performance and seamless running of the hotel, they should look at a few more intangible factors, such as staff service and personal customer feedback, rather than just a sample of it, rather than simply the data provided by TripAdvisor. Given that couples are the most common visitors to Las Vegas (as shown in table 3), it is suggested that the hotel provide more couple-specific discounts or amenity packages in order to attract audience-specific attention and, as a result, boost revenue and customer reach.

# BIBLIOGRAPHY:

Allen, M., 2017. Simple Descriptive Statistics - SAGE Research Methods. [online] Methods.sagepub.com. Available at: <https://methods.sagepub.com/reference/the-sage-encyclopedia-of-communication-research-methods/i13400.xml> [Accessed 23 March 2022].

Anon, (2021). 15 Mind-Boggling Las Vegas Revenue Statistics for 2021. [online] Available at: <https://playtoday.co/blog/las-vegas-revenue-statistics/> [Accessed 27 Mar. 2022].

Henrys, K., 2021. Role of Predictive Analytics in Business. 1st ed. [ebook] SSRN, p.a single page. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3829621> [Accessed 24 March 2022].

Lock, S., 2021. *Leading reasons for visiting Las Vegas in 2019*. [image] Available at: <https://www.statista.com/statistics/411638/main-reasons-for-visiting-las-vegas-us/> [Accessed 23 March 2022].

‌

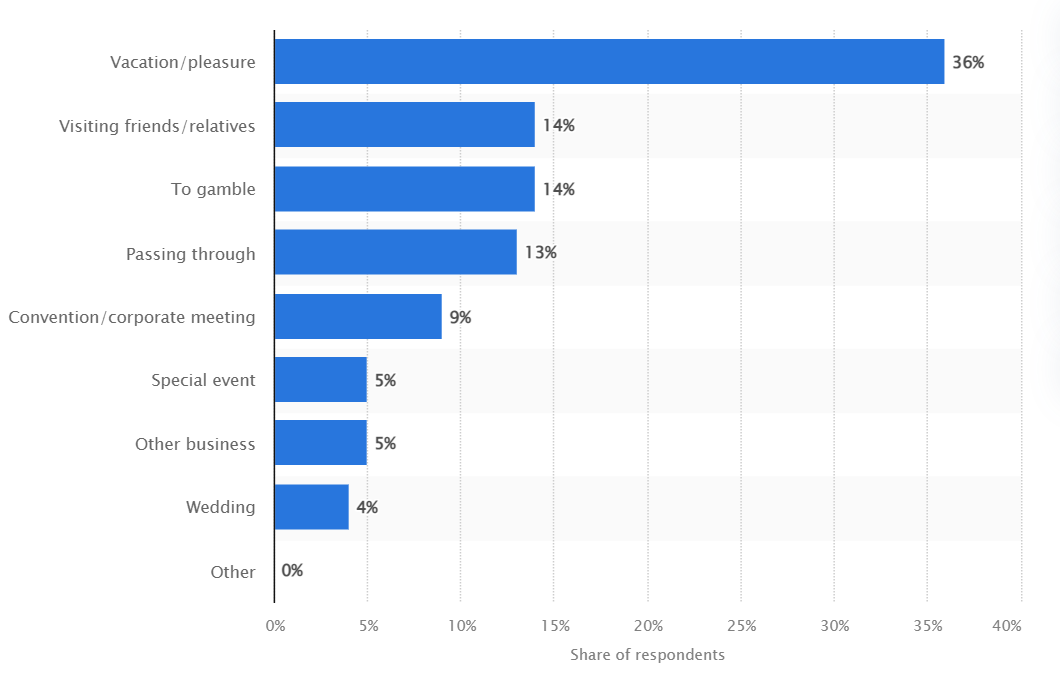
# APPENDICES:

Table 2

Table 3

Count of travellers based on travellers type

Table 4



(Lock,2021)

Reason to visit Las Vegas