

Making Expedia Great Again (MEGA)

Investigating how search filters affect the decisions of a rational consumer

Ishan Malhotra - 1007658553, Aarnav Chaudasama - 1007756395, Arnav Dey - 1008576140, Danish Ahmed Bombal - 1007733008 (GROUP 50)

March 31, 2022

Introduction

Through this presentation, we intent to peruse how consumer decision-making is influenced with changes in a multitude of filters while searching for different accommodations.

This in-depth investigation necessitated us to utilize variables such as the number of adults (*adult_count*), children (*child_count*), and infants (*infant_count*). Additionally, we used star ratings(*star_rating1*), number of reviews (*review_count1*), price bucket (*price_bucket1*) and whether a listing is discounted or not (*is_drr1*).

What are the average number of reviews generated for discounted and non-discounted listings? How well does star rating and review count predict the price of a listing? What is the majority consumer base of Expedia? These are some of the questions we hope to answer.

By the virtue of this detailed research study, Expedia will be able to comprehend how they can devise changes to the platform, make further improvements in Search Engine Optimization (SEO) and increase the overall appeal of their web page.

Data summary

Data Wrangling:

- Q1:
 - For Q1 we created two new variables: the first variable, *'total_no_of_kids'* is the sum of the *'infant_count'* and *'child_count'*, which indicates the total number of juveniles the individual is traveling with. The second variable created is *'traveling_with_kids'*, which indicates whether an individual is traveling with juveniles or not based on the *'total_no_of_kids'* count.
- Q2:
 - For this question, a new variable *'price_level'* was created taking price bucket 1 into consideration. If the price bucket was equal to 3 and below, then it is considered a low-price level. If the price bucket was equal to 4 and above, it is a high price level.
- Q3:
 - For Q3, we mutated a new variable named *'is_drr1_'* where we classified the listings whether they are at a discounted price or not and converted them in a more functionable format. [1
'Discounted', 0 *'Not discounted'*]

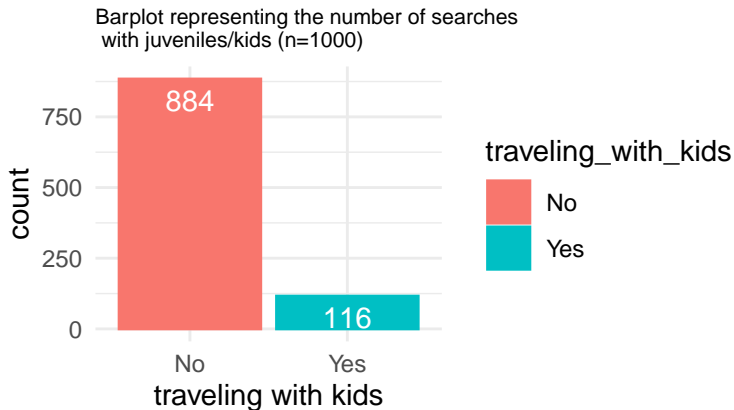
Question 1: Is the proportion of travelers who are seeking an accommodation with no kids accompanying them 90%?

Objectives

1. To identify, define and establish the majority customer base of Expedia.



(i) Visualization



(ii) Statistical Methods

Null Hypothesis

- Proportion of travelers on Expedia that are traveling without juveniles is equal to 90%

$$H_0 : p_{\text{traveling without children}} = 0.9$$

Alternative Hypothesis

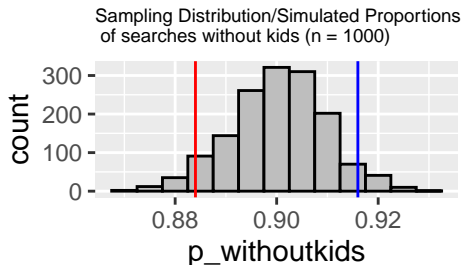
- Proportion of travelers on Expedia that are traveling without juveniles is not equal to 90%

$$H_1 : p_{\text{traveling without children}} \neq 0.9$$

(iii) Results and Interpretation

Since the significance level α (which determines the cut-off for how unusual/extreme the data is) is 0.05, and since the enumerated p-value is 0.09866667, we know that there is weak evidence against the null hypothesis i.e., we accept the null hypothesis.

1. 90 percent of customers are requesting for a property without children, which implies that it has a clear target customer it serves. Expedia can use this data to efficiently allocate their promotional budget in order to further increase their sales.



P-value = 0.09866667

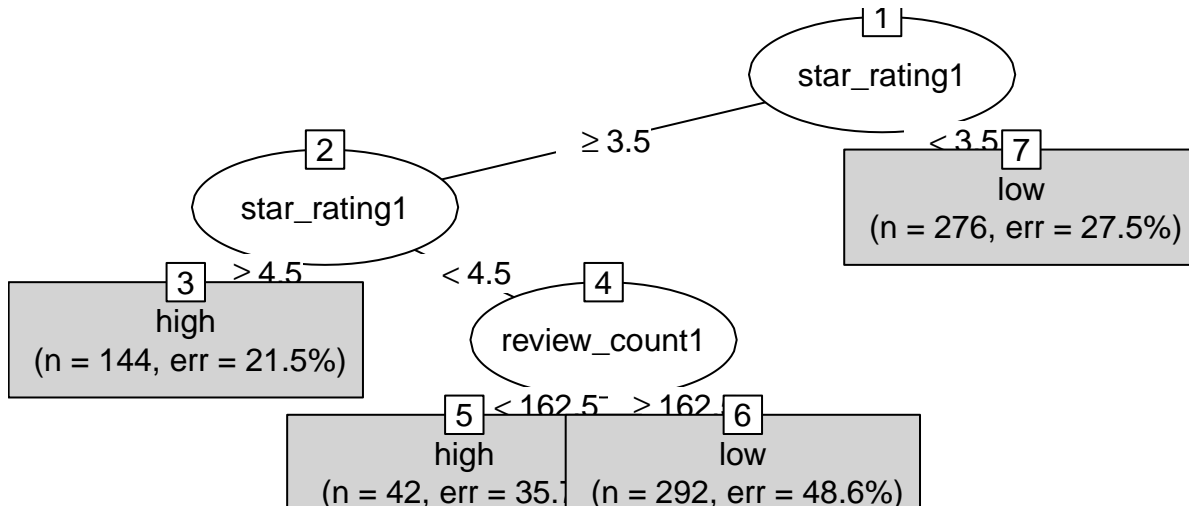
Question 2: How does review count and star rating affect the price level for the first listing?

Objectives

1. To increase the number of purchases by providing information about how review count and star rating affect the price level for the first listing



(i) Visualization



(ii) Statistical Methods

- Since price level is a categorical variable, we use a classification tree to determine whether review count and star ratings help predict price level. In this tree, the price level is the response variable whereas star rating and review count are the independent variables.
- The classification tree is modelled on training data which is 80% of all data. We then produce testing data that is 20% of all data and construct a confusion matrix in order to test the accuracy of the classification tree on the testing data.
- Additionally, we calculate the sensitivity of the classification tree (which is important for analysis purposes)

tree_pred	high	low
high	43	16
low	43	87

Accuracy: $(43 + 87)/(43 + 43 + 16 + 87) = 0.69$ **Sensitivity:** $(43)/(43 + 43) = 0.50$

(iii) Results and Interpretations

- The accuracy is 0.69 which implies that the prediction is correct 69% of the time. This means that this is a fairly suitable model for prediction
- Based on this analysis, we propose that new search filters are added namely review count and star rating that can be manipulated by the customer. Information should be given about which values of review count and star rating provide low prices and high prices. This should be done in order to maximize sales from those price-sensitive customers who look for low prices.
- The sensitivity is 50% which means that there is a 50-50 chance that the model predicts a listing to have a low price when it actually has a high price. If customers believe that prices which are supposed to be high are considered low on the site, they will psychologically convince themselves that expensive listings are actually cheap due to a lack of other options between low and high. Thus, Expedia can potentially earn more money.

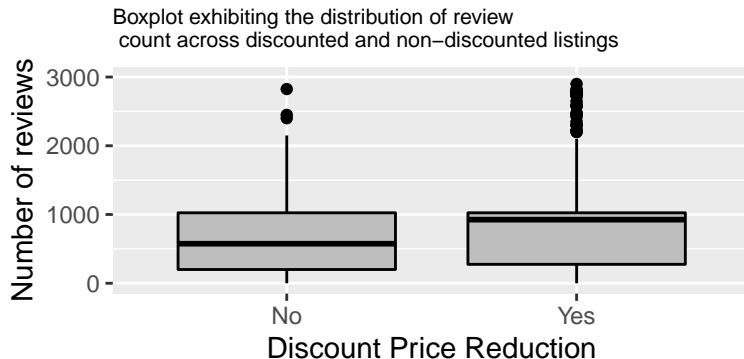
Q3: Should Expedia change their price discrimination (discount) strategies to surge the number of reviews on their listings?

Objectives

1. To assess whether the primary listing having a discount reduction enumerated price also has the highest number of reviews
2. To analyze and gauge the price marketing strategies of Expedia and ascertain if discount essentially plays a significant role in augmenting the number of reviews on a particular listing



(i) Visualization



(ii) Statistical Methods

Null Hypothesis

- There is no difference in the mean number of reviews for listings that are discounted and the listings that are not discounted.

$$H_0 : \mu_{discounted} - \mu_{not-discounted} = 0$$

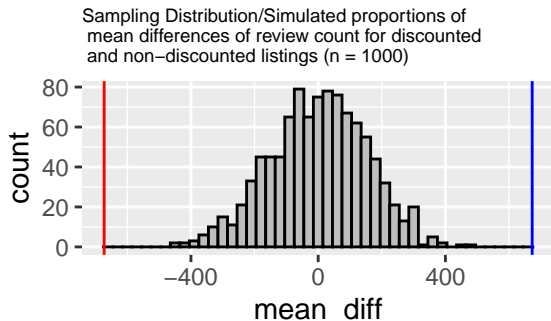
Alternative Hypothesis

- There is a difference in the mean number of reviews for listings that are discounted and the listings that are not discounted.

$$H_1 : \mu_{discounted} - \mu_{not-discounted} \neq 0$$

(iii) Results and Interpretations

1. The p-value is 0 i.e., there is very strong evidence against the null hypothesis that there is no difference in the mean number of reviews between discounted and non-discounted listings.
2. In compliance with the visualization, since there are more outliers in the boxplot for discounted listings than otherwise, the mean number of reviews is expected to be more substantial for the former.
3. Nevertheless, since there is an extremely slight and inappreciable difference between the IQR [Inter-Quartile range](Q3 - Q1) of the two boxplots, Expedia can implement superior ways to augment the number of reviews.



Test-Statistic: 672.8236

P-value: 0

Limitations

Q1

- During this period, the travel industry was extraordinarily impacted by the outbreak of Covid-19; hence, the data given to us is not archetypal (of the pre-pandemic era) as innumerable people would not risk going out with their children in such circumstances.
- There is also a probability of having a Type 2 error as we have accepted the null hypothesis. This insinuates that it is reasonably possible for the alternative hypothesis to be true.

Q2

- The thresholds set for a low price and a high price are arbitrarily determined. Possibly, what customers consider expensive and what we perceive they appraise expensive are quite different.
- The low sensitivity may lead to price-sensitive customers not booking listings because they recognize that the so-called discernible low prices are actually high.

Limitations (Cont.)

Q3

- There is a possibility that we might make a Type 1 error i.e., it is plausible that the mean number of reviews for both discounted and undiscounted listings are the same despite our pronouncement to reject the null hypothesis.
- Furthermore, there may be occurrences where a listing with a high number of reviews may also encompass a considerable number of substandard and poor feedback from the customers. Hence, it might not be an excellent decision to show such listings on the top.
- Though we have deduced that the mean number of reviews is not equal between discounted and non-discounted listings through the p-value, it is arduous to figure out the actual reason why they are not equal through this statistical approach. Perhaps, there may be numerous other factors that can remarkably influence the number of reviews including customer services, amenities available at the respective listing including parking, swimming pool etc. and more.

Conclusion

- For Q1, after the research, we were able to ratify and corroborate that the consumer base of Expedia majorly comprises of the people who travel without juveniles. Through this information, Expedia should now utilize price discriminating strategies on clientele groups like couples; consequently, increasing their overall sales.
- For Q2, the addition of filters can help proliferate the number of purchases for Expedia. Price-sensitive customers can be guided to listings suitable to their preferences who did not have the option of using these filters. In the investigation, we ascertained that the patrons may not have booked a listing because they could not find a suitable price in the recommendations given to them. This is a loss for Expedia and the consumer as a result of insufficient website maneuverability.
- For Q3, we conclude that it is indispensable for Expedia to give out more discounts to increase the number of sales (from Q1 through price discrimination strategies), thus increasing the number of reviews. A research study has demonstrated that 67% of the consumers are eager to write a review if they are proffered a discount on a stipulated good or a service (Pitman, 2022).

References and Acknowledgements

The creators of this project would like to thank Nayan (TA) for helping us with the development of this project, permitting us to more readily comprehend and change questions, address concerns and limits, and how to make the end result

- Pitman, J. (2022, January 26). Customer Review Trends 2022. Retrieved from Bright Local: <https://www.brightlocal.com/research/local-consumer-review-survey/#:~:text=67%25%20of%20consumers%20will%20consider,leave%20reviews%20for%20positive%20experiences.>

