# Using Customer Data to Inform Incentive Deployment

## Buckeye Resorts, Inc.

*Data Science Team: Xavier Johnson, Brian Wisniewski, Matthew Romanko*

# Problem discussion

- Buckeye Resorts, Inc. (BRI) loses $500 for every cancelation on average
- A pilot program aimed at reducing cancellations found $100 discounts were effective 30% of the time
- BRI would like to understand if using customer data could help predict the likelihood of a particular reservation to be canceled
- If the data is sufficient for developing an accurate model, BRI could improve on the 30% efficacy and reduce the overall cost of the incentive program

# Customer reservation data

- 119,391 customer reservations
- Three years of data (2015-2017)
- The Original data contained 32 attributes for each customer reservation

# Solutions approach

**Goal:** Utilize existing customer data to predict the likelihood of a particular reservation being cancel

- Initial exploratory data analysis
- Customer attributes chosen for model
  - Predictor variable
    - **is_canceled**: indication of a customer reservation being canceled
  - Explanatory variables
    - **kidsToAdults**: ratio of the number of kids to adults (kids are a sum of babies and children)
    - **cancelRate**: previous_bookings_not_canceled - previous_bookings_canceled / total number of bookings
    - **correctRoom**: logical where reserved_room_type == assigned_room_type

# Solutions approach

- Develop cost matrix for evaluating modeling benefit

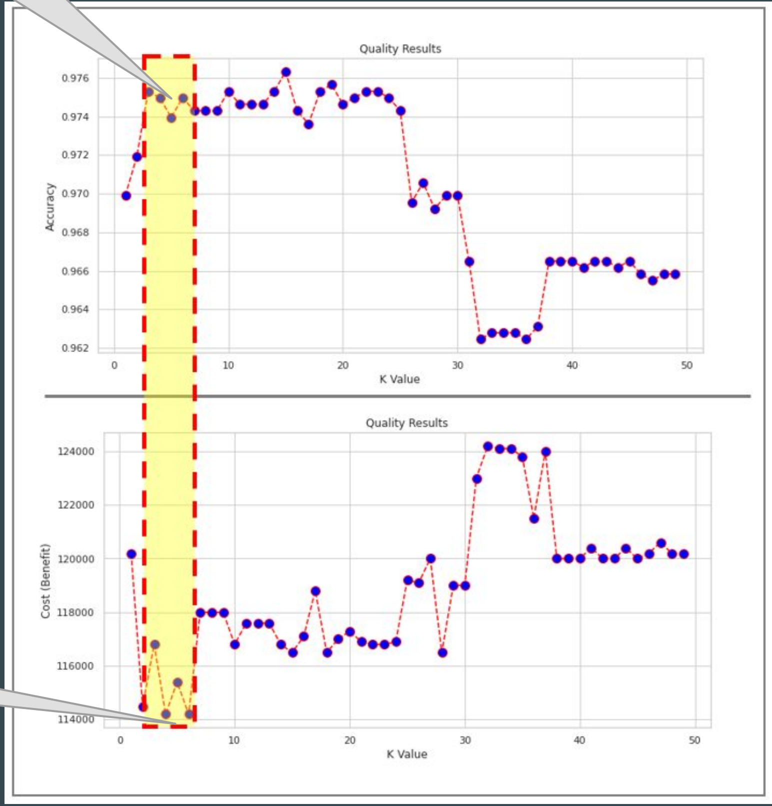
| | PREDICTED VALUES | | |
|---|---|---|---|
| | canceled | didn't cancel | |
| canceled | 100 | 500 | TRUE VALUES |
| didn't cancel | 100 | 0 | |

- Implement KNN model for prediction

# Model Results

❖ Model results predict high accuracy can be obtained from this model at low cost using the variables chosen with this approach

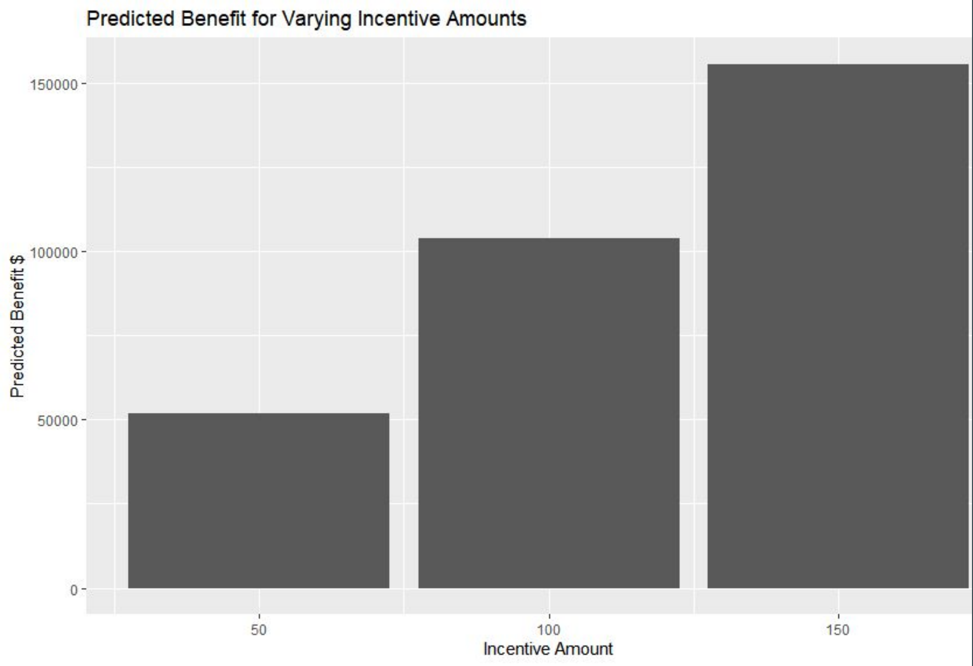
High Accuracy




Low Cost

# Recommendation & Discussion

| | |
|---|---|
| Based on the analysis results, we recommend BOD moves forward with the approach. | |
| Current data is limited. | • We do not have real-world information on how changing the incentive amount impacts the bottom line. |
| We propose an amended approach to address this. | • It Involves trialing different incentive amounts ($50, $100, $150) for 1 year so that we can determine the most profitable incentive amount. |
| We believe this amended approach will make the problem more tractable. | • It will provide real world data to supplement our model predictions. |
| According to our confusion matrix predictions, we expect these benefits from the varying incentive amounts: | • $50: $51,900 total, and per instance ~$17<br>• $100: $103,800 total, and per instance ~$35.<br>• $150: $155,700 total, and per instance ~$52. |



# Conclusion

- In conclusion, we constructed a model with the predictor variable is_canceled and the explanatory variables **kidsToAdults, cancelRate, and correctRoom.**
- **With this model, we are confident that we are able to calculate the "at-risk" reservations that are made.**
- We also concluded that implementing some amount of compensation would in fact beneficial to following through with reservations.
- The amount of compensation is not certain, as we feel as there is not enough data to make an amount with conviction.
  - We suggest doing a couple of trials with varying amounts of compensation during this pilot period in effort to maximize the profit from each reservation that follows through.