**Hotel Booking Demand**

Problem Addressed

Hotel Booking data is a booking summary for a hotel and a resort in the same city. Various specs of booking were provided including arrival, length of stay, number of guests, room type, special requests, price, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. With a total of 32 variables and 110K+ observations, it proved a viable dataset for our research. We build models to predict: the number of special requests, the best time of year to book a hotel/resort room and, the optimal length of stay for the best daily effective rate.

The choice of suitable analytical technique(s) questions being addressed.

Our approach to the data was based on the research questions posed during the exploratory analysis. We determined that in order to answer the questions, we needed to comprehend the data on hand in detail.

First, we clustered the data. The original dataset consisted of 32 variables. Through clustering, we sought to group variables and find patterns among different groups. We began this process by cleaning the data. This was done by eliminating missing values as they were only four, formatting variables - for example, converting integers to numeric, and removing outliers. Next, to reduce the number of variables, for a more focused analysis, we determined the variables that were irrelevant to our purpose and removed them from the dataset. These variables primarily had ‘NULL’ observations.

With numeric variables only in the cleaned dataset, we performed a k-means clustering. Next, as a result of clustering, arbitrarily removing variables did not feel right. Hence, we turned to PCA - Principal Component Analysis, for a second check. This analysis allowed us to further trim the variables in the data, thereby increasing the interpretability of it.

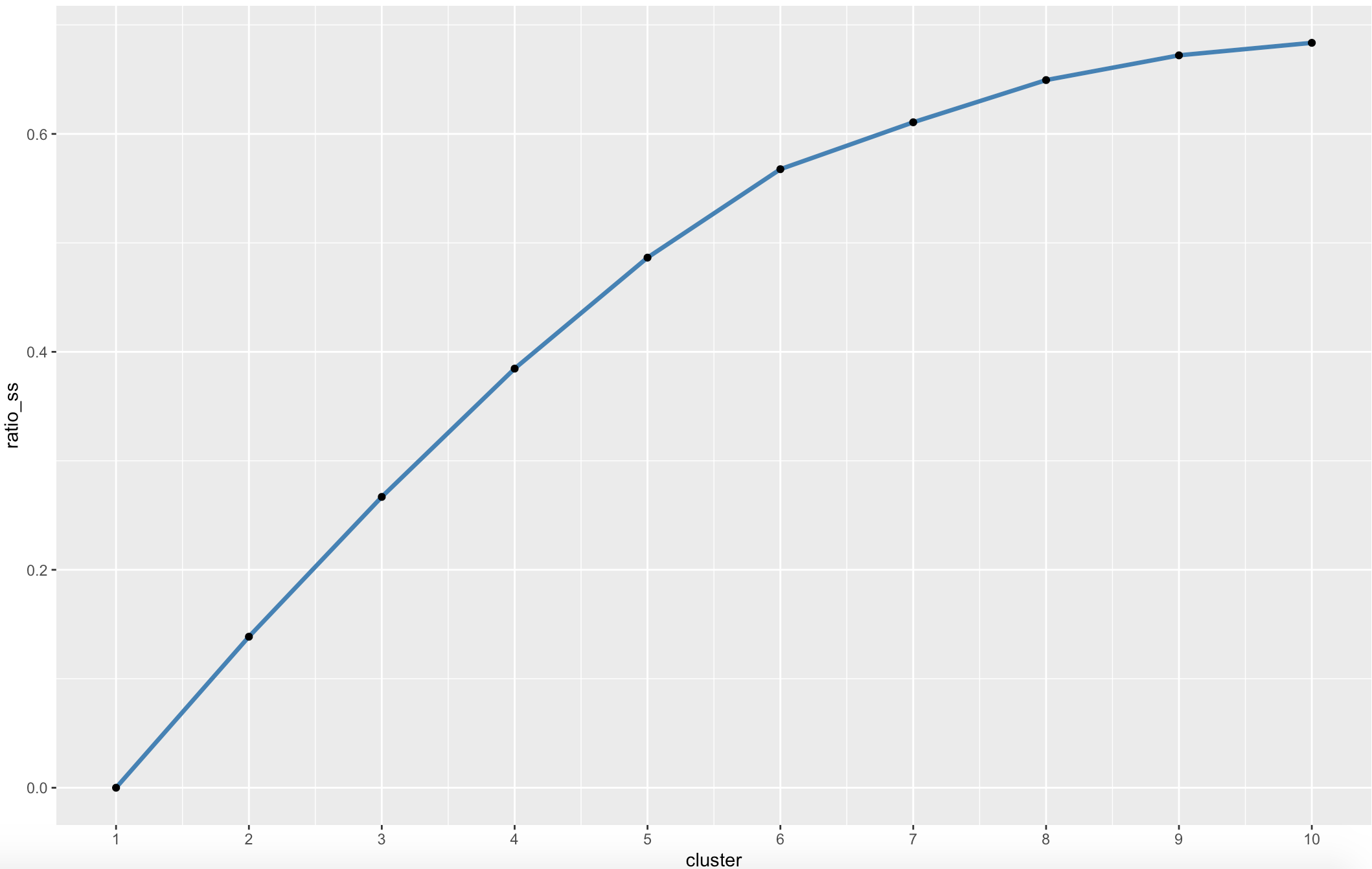
While conducting the k-means clustering again, we turned to the sum of square plot and ratio plot to determine the optimal number of clusters. Once we got comfortable with that, we answered our research questions: predict the number of special requests, best time of the year to book and optimal length of stay for the best effective daily rate.

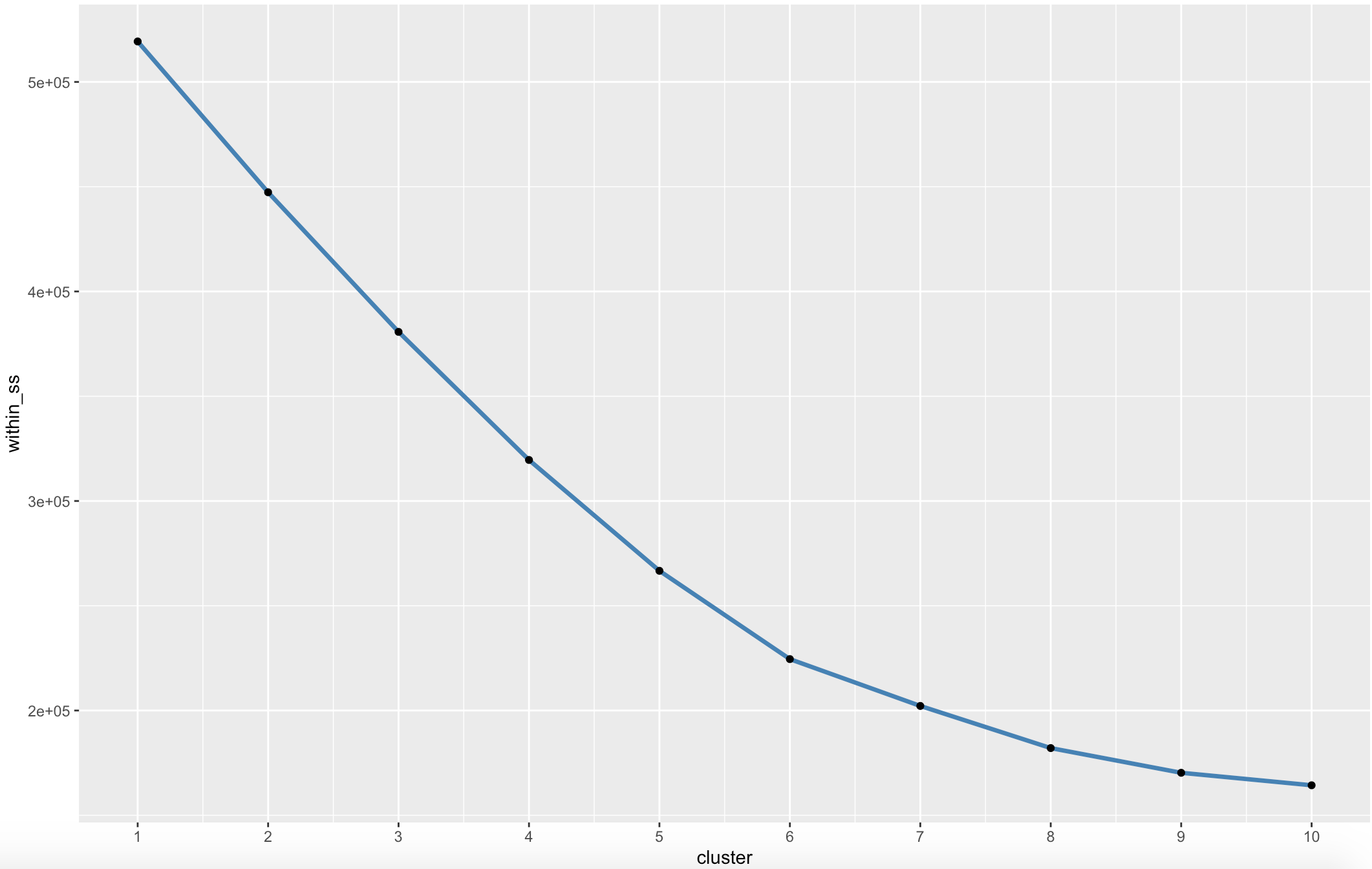
When it came to predictions, we also had our comings and goings with some trial and error. We started with the traditional linear regression but later realized that predicting using trees lowered the SSE - the sum of squared estimate of errors. The tree model was more accurate than the linear regression model.

Results from the Analyses run

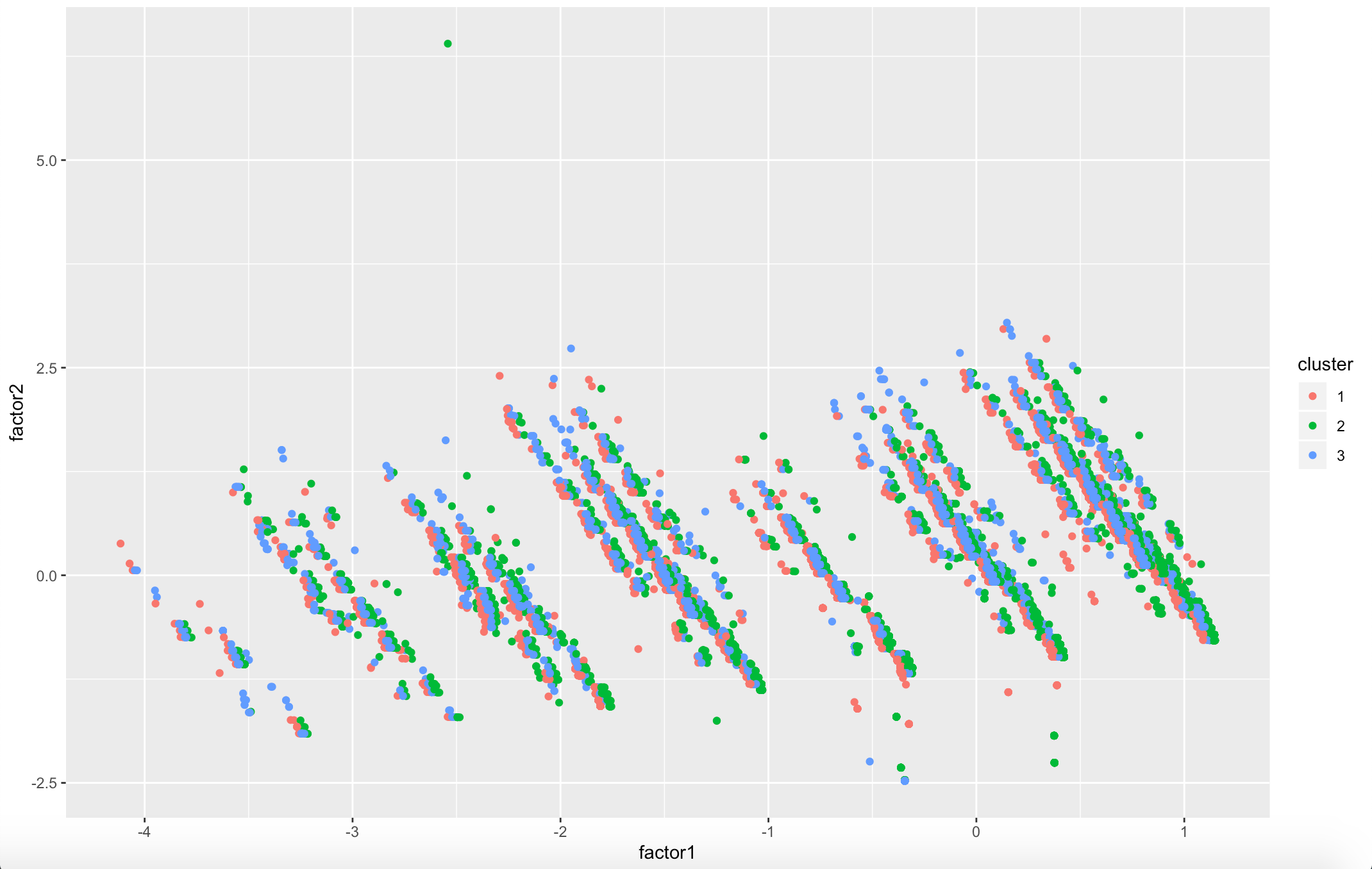
1. Predict the number of special requests made by customers

As explained above, based on our research, we found that we could cluster the data into 3 groups. To come to this conclusion, we had plotted the ratio plot against the total within sum of squares plot (as shown in the graphs below).



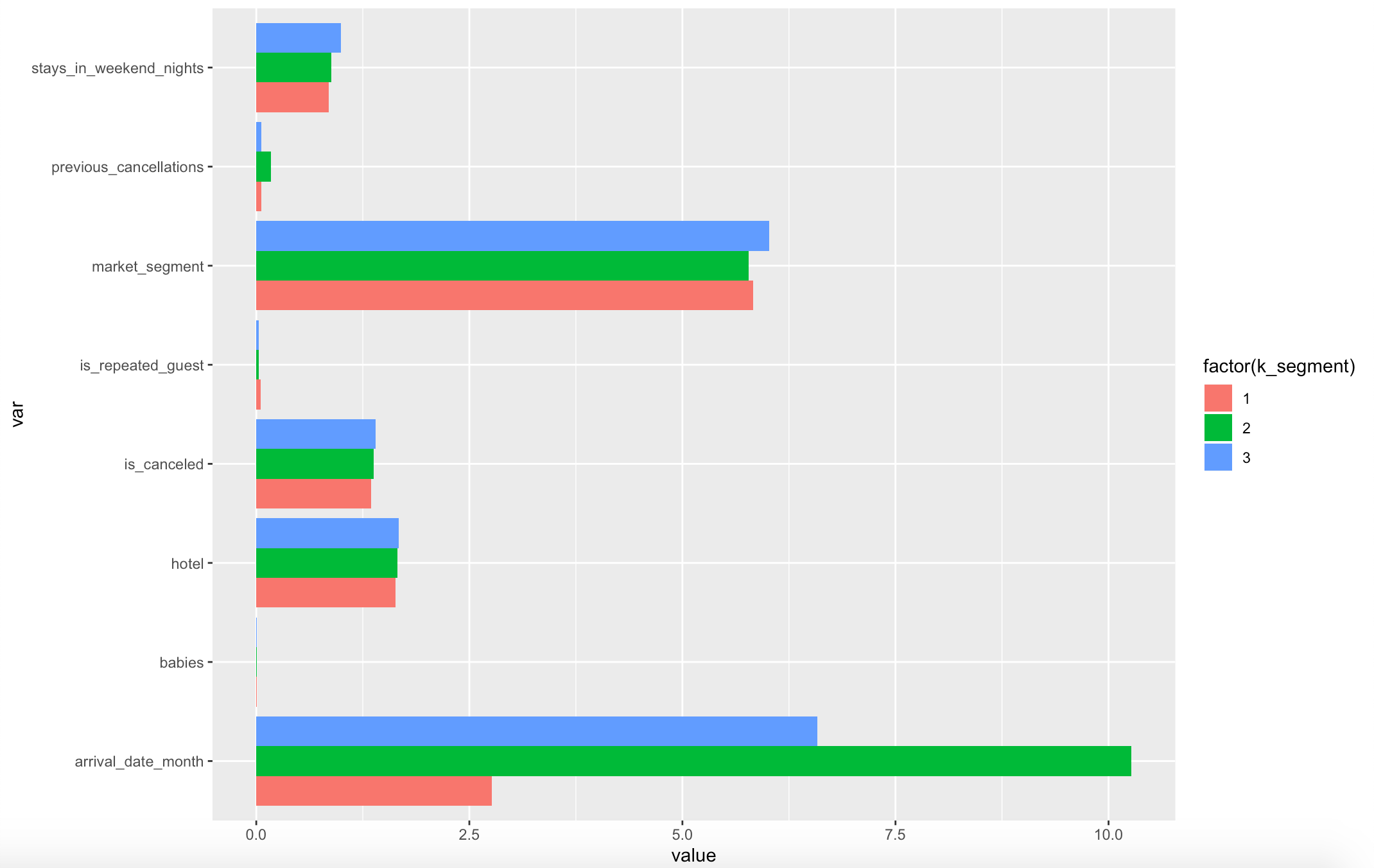


After determining the number of clusters, we used k-means to perform clustering and plotted the results.



In order to learn more about the different clusters we ran a profile on them. This way we could inspect the characteristics of each cluster and how they differed amongst each-other.

We believe this can be extremely useful for hotel owners to help identify what different customer segments value to deliver more targeted offerings.



The chart above allowed us to figure out how each group is built. While arrival month and market segment - that is whether for example clients booked by a Travel Agent or Tour Operator, are the key factors in these clusters, we can observe the importance of variables on the x-axis is different in all three groups. For example, clearly Group 1 is not that heavy in arrival month, while Group 2 is the heaviest in arrival month. Finally, Group 3, on the other hand, looks more evenly spread, between the two features.

Our goal with this first research question was to try to predict the number of special requests clients could ask for when booking. So, after splitting the data and training it, we used linear regression and tree models to predict the number of total special requests by cluster. We used both methods as we wanted to see which one yielded the lower SSE. Finally, we chose the predictions made using the tree model.

We arrived at the following conclusions through the analysis:

* Originally, the hotel bookings dataset contained special requests that ranged from 0 to 5. Evaluating 3, 4 or 5 special requests seemed a bit out-of-scope. Especially considering that we did not have insights through the data on guests with disabilities, and the fact that whether or not each booking had requested a parking space, was already accounted for separately.
* We were able to observe that special requests in our prediction model ranged from 0 to 2. In order to be better prepared, the hotel would want to know the month in which the most special requests would be required and the market segment from which those requests would come from; hence, we evaluated this as well.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Hotel | Predicted Special Requests | Month | Segment |
| 1 | City | 1 | August | Direct |
| 1 | Resort | 1 | July | Offline TA/TO\* |
| 2 | City | 2 | August | Complementary |
| 2 | Resort | 2 | May | Offline TA/TO\* |
| 3 | City | 1 | July | Online TA\*\* |
| 3 | Resort | - | - | - |

\*Offline TA/TO: offline travel agent or tour operator

\*\* Online TA: online travel agent

1. Predict the best time of year to book a hotel/resort room

The approach to this question follows a similar path to the one described for answering our first research question. Using the same clusters and the same cleaned dataset, we embarked on trying to determine the best time of the year to book a hotel/resort room.

After splitting the data and training it, before moving to the test dataset, we ran both linear regression and tree models. The objective here was the same: to find the model with the lower SSE. It turned out to be trees again.

Once we got our prediction, we ordered, filtered and grouped the dataset. Next, we sought the smallest ADR to find the best time of the year to make a reservation.

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Hotel | Predicted ADR | Month |
| 1 | City | - | - |
| 1 | Resort | 55.3 | January |
| 2 | City | 64.7 | June |
| 2 | Resort | 44.9 | March |
| 3 | City | 80.9 | January |
| 3 | Resort | - | - |

1. Predict the optimal length of stay for best ADR

Since we already had predicted the ADR to answer our second research question, we thought it would be a good idea to try to determine what would be the optimal length of stay.

To do this, we calculated the median for stays in weeknights and weekend nights per cluster. This way, we would be able to find the number of nights that people stay in city or resort hotels. Following that, we ordered, filtered and grouped the data, in order to get the ADR that matched those medians.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Hotel | Predicted ADR | ALS\* | Weekend / Weekdays |
| 1 | City | - | - | Weekdays |
| 1 | Resort | 55.3 | 3 | Weekdays |
| 1 | City | - | - | Weekend |
| 1 | Resort | 55.3 | 1 | Weekend |
| 2 | City | 64.7 | 1 | Weekdays |
| 2 | Resort | 44.9 | 1 | Weekdays |
| 2 | City | - | - | Weekends |
| 2 | Resort | 44.9 | 1 | Weekends |
| 3 | City | 80.9 | 2 | Weekdays |
| 3 | Resort | - | - | Weekdays |
| 3 | City | 80.9 | 1 | Weekends |
| 3 | Resort | - | - | Weekends |

Conclusions and Recommendations

Hotels seem to have more special requests during the months of summer, regardless if it is a city hotel or a resort hotel. That seems to be logical as those months are when most people tend to leave for vacations and hotel vacancies tend to be the least. One curious fact though, is that Resort Hotels for Group 2 would expect the peak of special requests in May, before summer vacations begin.

The best time of the year to book a hotel for the cheapest rates is during the cold weather season. Depending on the city and area where these hotels may be located, winter could be the down season for them if there are no events or attractions. This also goes with the fact that the highest number of special requests take place during warmer months, when it is expected that the majority of people would be taking vacations.

Finally, it looks like the best way to get by with staying at hotels at the best possible rate, is to keep it simple - booking for one night in most cases, seems to help the guests get the lowest rate. This might be because of the promotions hotels put out there in order to attract customers.