Name: Dikshant Joshi

Lab 2 : Causal Forest Revisited

Question 1: **In a table, list the variables you decided to include in your models and explain why.**

I decided to chose the variables that were required in Q4 which were bedrooms, accommodates, minimum_nights, review_scores_rating, room_type, state, city and price, high_booking(As outcome variable) and Treatment(*Coded Airbnbs in and after Nov. 2019 as 1; and otherwise 0)*.

Also these are the variables that may influence the booking rates of Airbnb as follows

- For Price as the Outcome Variable

Variables Used	Importance
Bedrooms	➢ Having a larger number of bedrooms can often be seen as a desirable feature by guests, as it provides more space and privacy for groups traveling together. As a result, listings with more bedrooms may be more likely to attract bookings from guests looking for larger accommodations, especially for family or group trips.
Accommodates	➢ Accommodates refers to the maximum number of guests a listing can accommodate, which can be a critical factor for guests looking for a suitable place to stay. A listing with a high number of accommodates can attract larger groups of guests, which can increase the likelihood of high bookings.
Minimum Nights	➢ The minimum nights requirement can be a factor in guest decision-making, as some guests may prefer more flexibility with their travel plans. A higher minimum nights requirement may deter some potential guests from booking, while a lower requirement may make the listing more attractive to those looking for a shorter stay.
Review_Scores_rating	➢ A higher review_scores_rating is generally associated with a higher likelihood of high bookings, as it suggests that previous guests have had a positive experience and are more likely to recommend the listing to others. On the other hand, a lower review_scores_rating may suggest that there are issues with the listing, which can negatively impact the likelihood of high bookings.
Room_type	➢ Room_type variable can be an important factor to consider when predicting the likelihood of high bookings in an Airbnb listing, it should be evaluated alongside other relevant factors to obtain a more accurate prediction which is why location and price variable will additionally help.
State	➢ The State variable represents the location of the listing and can be important in determining the likelihood of high bookings. For example, a listing in a popular tourist destination such as Hawaii or California may see a higher demand for bookings, resulting in a higher likelihood of high bookings.

Name: Dikshant Joshi

City	➢ Same like state variable city can represents the location of Airbnb and can be important in determining the likelihood of high bookings.
Price	➢ Price variable can be important in determining the target market for the listing. A listing with a high price may attract guests looking for luxury or high-end accommodations, while a listing with a lower price may attract budget-conscious travelers or those seeking a more affordable option.
Treatment	In November, 2019, after a Halloween night shooting at a "party house" in Orinda, California, Airbnb pushed a number of safety measures and rules.
High_booking	The outcome variable

Question 2: **Estimate the conditional average causal effect on the booking rate (and standard error)**

- **Use average_treatment_effect() to calculate the treatment effect**
- **target.sample would depend on the effect you will be seeking**
- **You'll have to set this to "overlap" if you receive a warning or NaN**
- **Set method argument to AIPW (Augmented Inverse-Propensity Weighting)**
- **Use subset argument to answer questions on treatment heterogeneity**

```
#Calculate Average Treatment Effect on our outcome variable
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"))
```

```
##    estimate     std.err
## 0.066363372 0.009793245
```

Question 3: **The causal effect you found above is a doubly robust estimate of the average treatment effect. But you did not split the data and set aside a completely isolated test set to start with. To confirm your results are robust, split your data and use predict() to calculate individual effects in the test set, and take an average of the individual effects to find the average treatment effect. Is the new effect consistent with the effect you found above?**

```
df1.test <-
  predict(cf_split, X2.test) %>%
  bind_cols(df1.test)

df1.test%>%
  summarise(ATE = mean(predictions),se=std.error(predict(cf_split, X2.test)))
```

```
##          ATE           se
## 1 0.06164293 0.0005811336
```

Name: Dikshant Joshi

Question 4: **See the hints in Q2 and calculate the heterogeneity in the average treatment effect for:**

a.Airbnbs with few or many bedrooms (Use bedrooms)

```
#Heterogeneity of ATE for Airbnbs with few or many bedrooms
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=X[,1]>1)
```

```
##   estimate    std.err
## 0.04607106 0.01415550
```

b.Airbnbs that accommodate more or less people (Use accommodates)

```
#Heterogeneity of ATE for Airbnbs that accommodate more or less people
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,2]>10 | X[,2]<5))
```

```
##   estimate    std.err
## 0.07193539 0.01194974
```

c.Low or high minimum nights requirements (Use minimum_nights)

```
#Heterogeneity of ATE for Airbnbs which have Low or high minimum nights requirements
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,3]>466 | X[,3]<233
```

```
##    estimate     std.err
## 0.066197817 0.009802044
```

d.Higher or lower rated Airbnbs (Use review_scores_rating)

```
#Heterogeneity of ATE for Higher or lower rated Airbnbs
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,4]>66 | X[,4]<33))
```

```
##    estimate     std.err
## 0.066122149 0.009815698
```

e.Whether Airbnbs are the entire house or not (Use room_type)

```
#Heterogeneity of ATE Whether Airbnbs are the entire house or not
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,5]==1))
```

```
##   estimate    std.err
## 0.06659807 0.01116819
```

Name: Dikshant Joshi

f.For three states and cities in different parts of the country (Use state and city)

State:

```
#Heterogeneity of ATE For three states and cities in different parts of the country
#State CO
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,9]==1))
```

```
##  estimate    std.err
## 0.1157297 0.0418297
```

```
#State DC
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,10]==1))
```

```
##     estimate      std.err
## -0.01467874   0.04727981
```

```
#State FL
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,11]==1))
```

```
##   estimate    std.err
## 0.02845460 0.02040599
```

City:

```
#City Austin
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,27]==1))
```

```
##   estimate    std.err
## 0.09007115 0.05856985
```

```
#City Boston
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,28]==1))
```

```
##     estimate     std.err
## -0.13527132  0.08590961
```

```
#City Chicago
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,31]==1))
```

```
##   estimate    std.err
## 0.03434872 0.04926678
```

g.One source of heterogeneity you think is interesting to check (Dealer's choice!)

```
#Heterogeneity of ATE for One source of heterogeneity you think is interesting to check(price>500)
average_treatment_effect(cf,target.sample=c("overlap"),method = c("AIPW"),subset=(X[,54]>500))
```

```
##   estimate    std.err
## 0.07174725 0.04804818
```

Name: Dikshant Joshi

Question 5: **In a single paragraph, interpret, explain, and discuss your findings. What do you think is your most interesting finding and why? If it is counterintuitive, how do you explain it?In addition, use bullet points at the end of your write-up to list your key takeaways and lessons learned from the analysis. Feel free to connect the dots with the previous lab.**

The overall effect of our treatment variable over outcome variable is positive which tells that after the rules was released the booking rate increased . Most interesting finding was for one of the city Boston and state DC where I got the effect size negative but the standard error is huge so not sure if the effect is significant. Key takeaway from this lab is how important it is to use cluster while making causal forest working with panel data.