

Booking Trend Analysis

Summary of the Booking Data

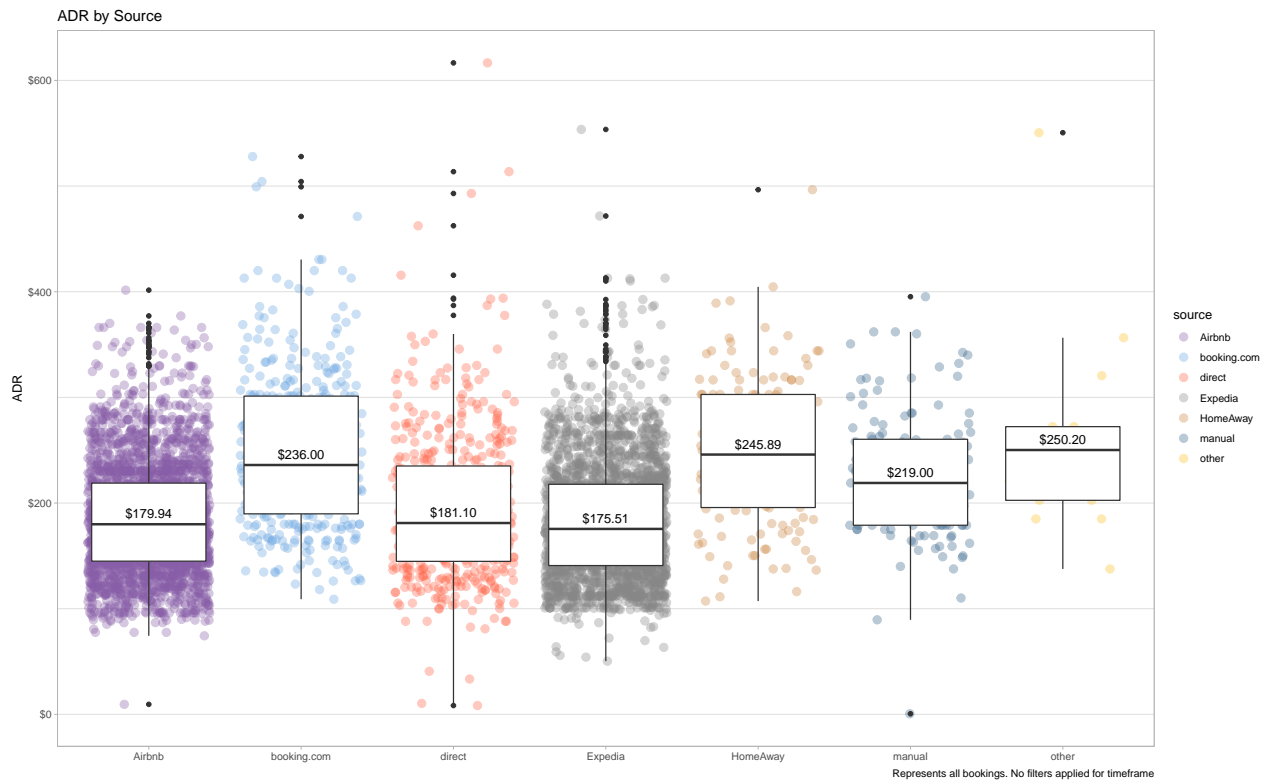
Summary statistics of data show booking sources and the number of bookings that came through each. Note the large outliers in some of the max and min values. The records warrant further investigation to understand if these outliers are input errors and should be corrected or discarded. The data has not been filtered by any timeframe, and represents all Domicile confirmed bookings.

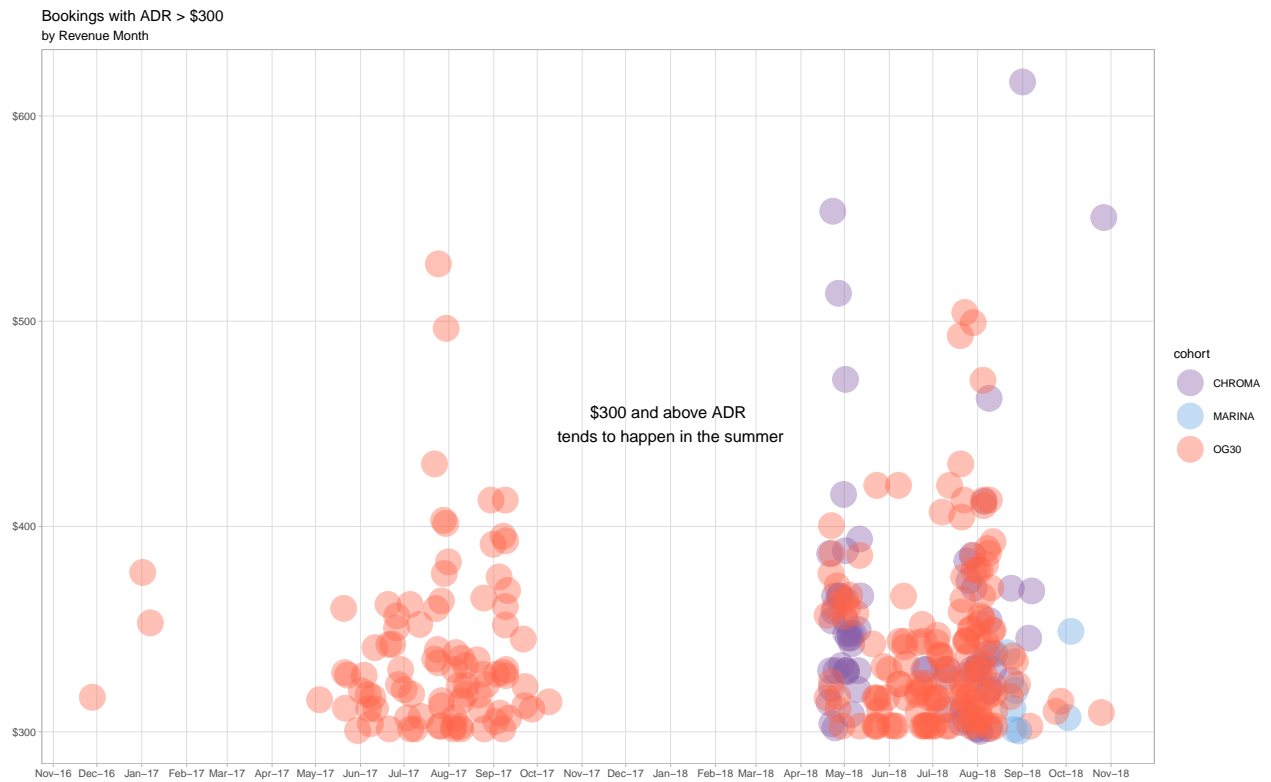
##	source	ADR	num_nights	days_in_advance
##	Airbnb	:2615	Min. : 0.55	Min. : -293.00
##	booking.com	:364	1st Qu.:146.02	1st Qu.: 2.00
##	direct	:467	Median :182.01	Median : 10.00
##	Expedia	:2354	Mean :191.70	Mean : 16.81
##	HomeAway	:140	3rd Qu.:227.69	3rd Qu.: 25.00
##	manual	:135	Max. :616.56	Max. : 489.00
##	other	:14		

Bookings by source

Boxplots show the summary distribution of data: min, max, median and quartiles. The box encloses values that represent 50% of the data, bounded by the 25th and 75th quartiles. The “whisker” lines extend to min and max values. Outliers to the data are shown with black dots. Colored dots behind the plot are individual bookings, to give perspective on the number of values shown in the distribution.

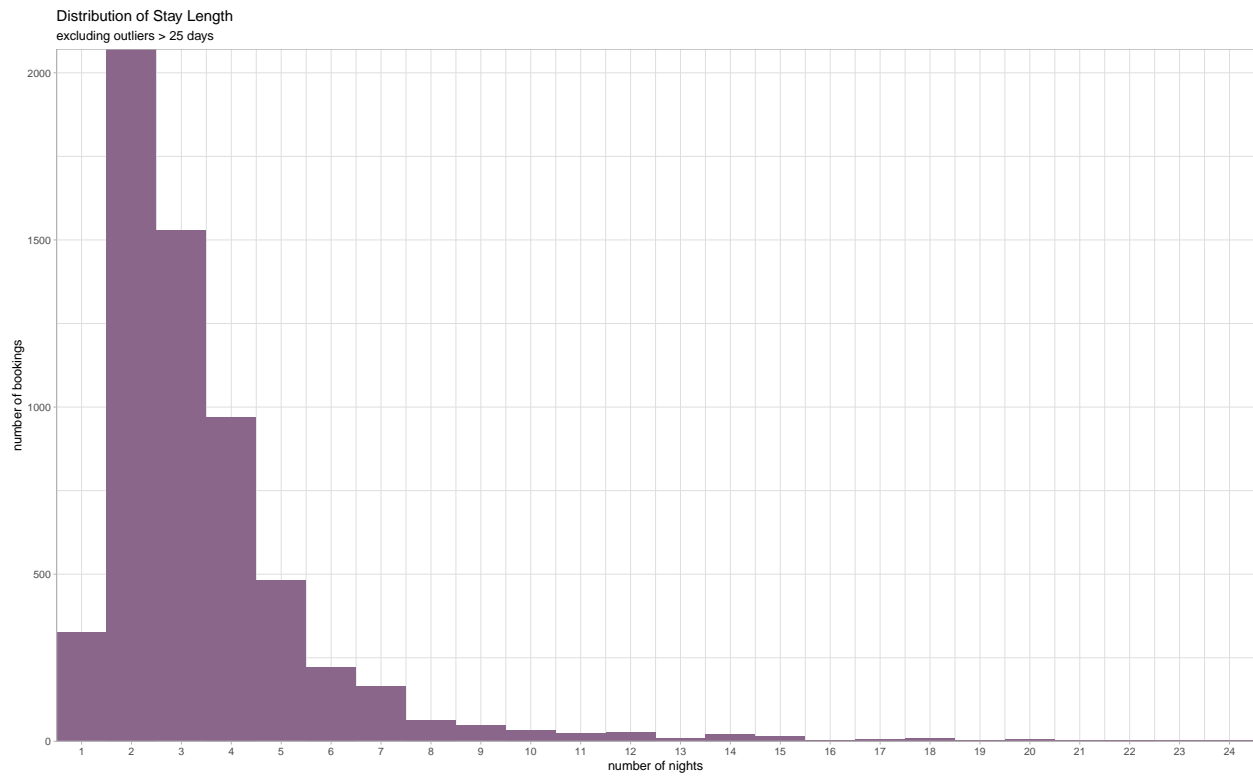
Airbnb and Expedia represent the largest venues for bookings, and also the lowest median ADR. Some of the lower volume booking venues seem to provide bookings at higher rates. It would be interesting to drill down on the demographic of the higher rated bookings. Do they occur in specific months, or are the customers from any particular region?





Length of Stay

The histogram shows distribution of the raw booking data by length of stay. The data had large outliers. These were filtered out to provide better detail on the majority of the data.



Trend by Day of Check In

Is there any trend by day of the week? Do check-ins that begin on different days have different average durations? We can see that there are slightly more check-ins on Friday, with shorter durations of stay. We can see that Fridays are more likely to have only 2 nights, consistent with weekend travelers. Thursday check-ins are more likely to stay into the weekend. And check-ins on Sunday have the longest average duration of stay. Median values were used instead of average to reduce the impact of large outliers on the average.



Booking in Advance

Does how far ahead a booking is made have any impact on price? This plot shows the distribution of bookings by ADR and the number of days ahead of check-in that the booking was made. A linear smoothing model was applied, with standard error shown in light gray. It doesn't appear that there's a strong trend, and the large outliers probably have impact on the slope of the regression line.

Price Regression Model
Is price impacted by how far in advance the booking is made?

