

THOMAS CARTWRIGHT

EXPEDIA – CASE STUDY

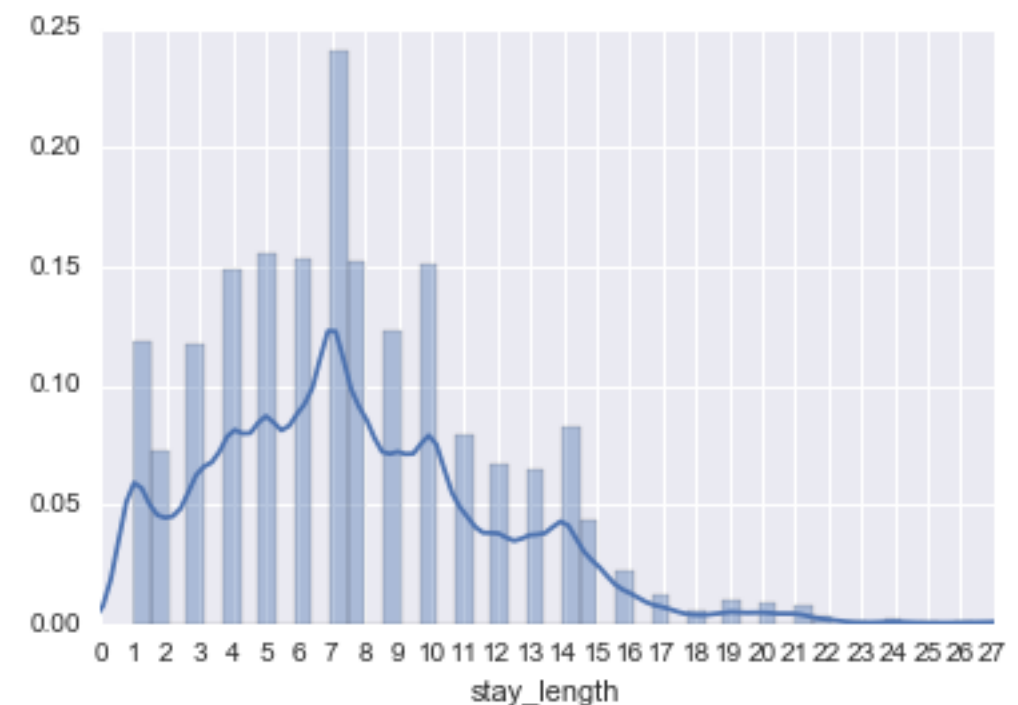
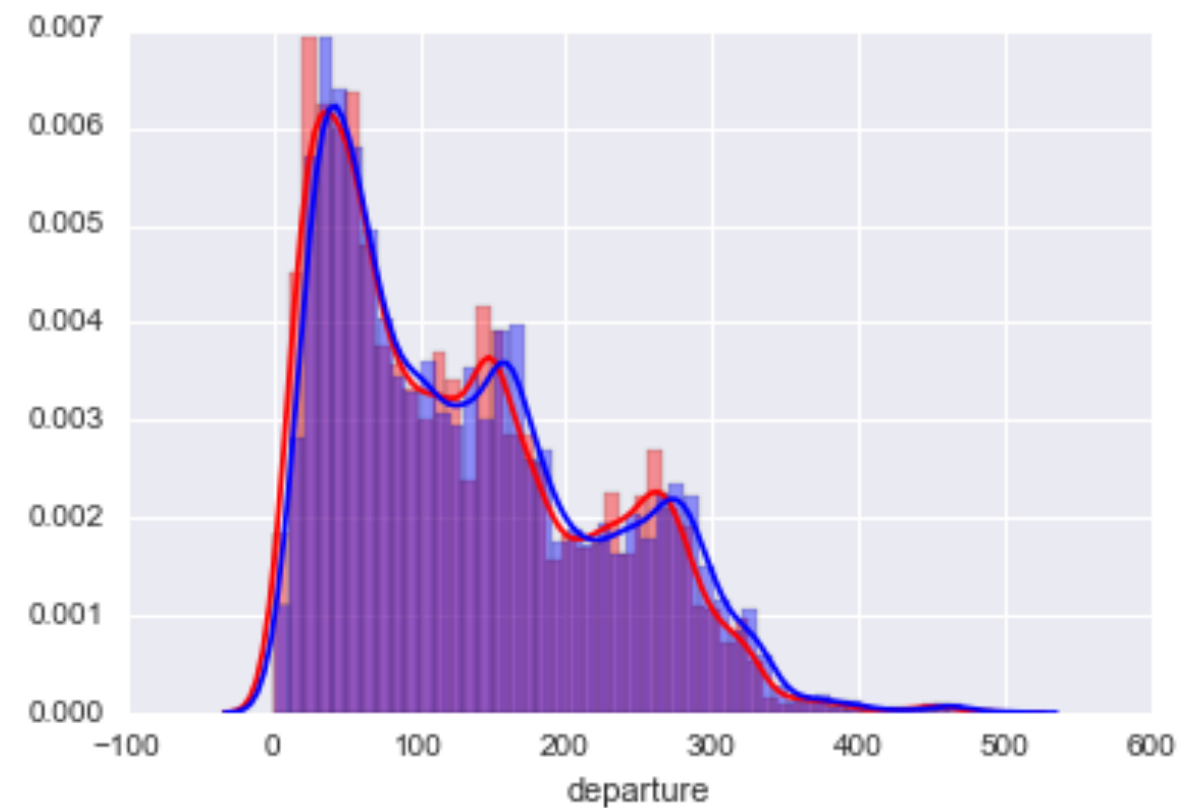
AIM OF PRESENTATION

- ▶ Look at distributions and a summary of the data
- ▶ Look at any correlations (or lack of) within the data
- ▶ Discuss a model for predicting hotel price based on features and user input
- ▶ Discuss possible improvements with the data source

SUMMARY OF DATA

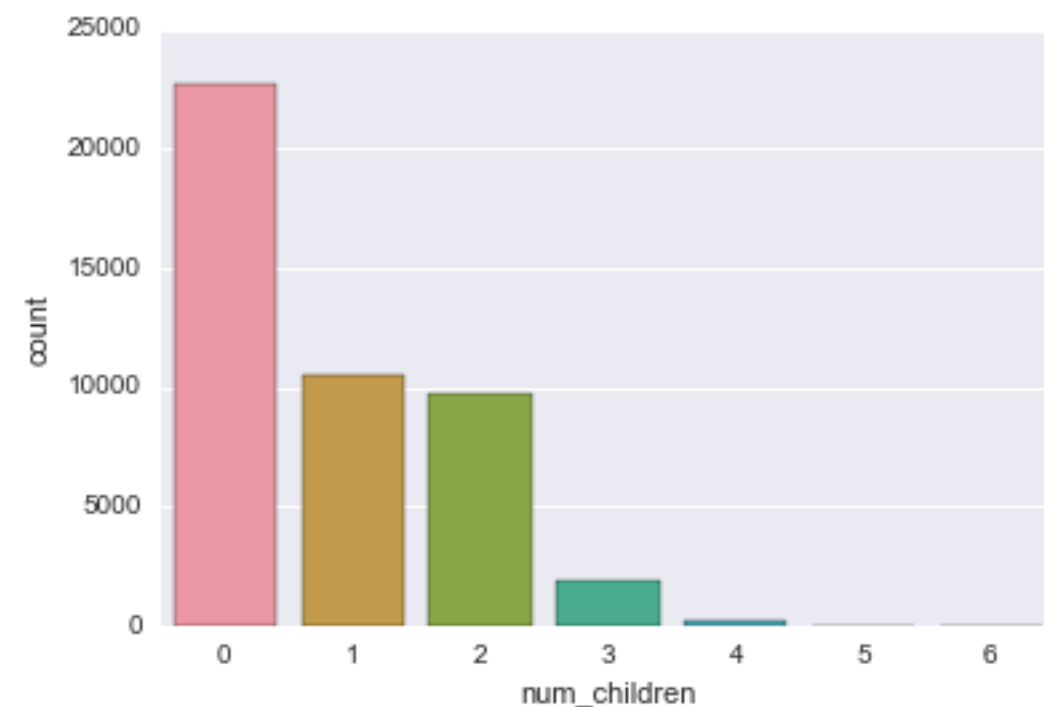
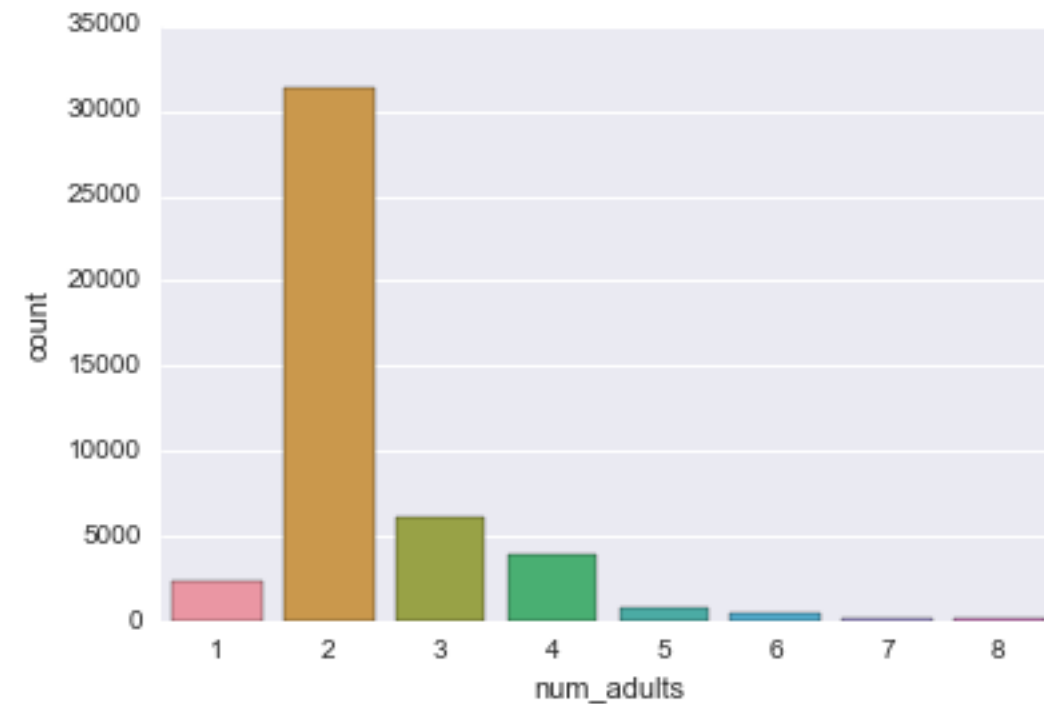
HOLIDAY DATES

- Most popular dates to depart are during January and then the summer (though this is skewed because our data is only the bookings made in the first month of the year)
- Most popular length of stay is 7 days with local maximums around 10 and 14 days.



FINDING THE PERFECT HOTEL

- We can see from graphing that by far the most popular enquiry is for 2 adults
- We can also see from graphing that the most popular enquiry is for just adults and no children
- We also found that on average 3 hotels are viewed in each search session



CORRELATIONS

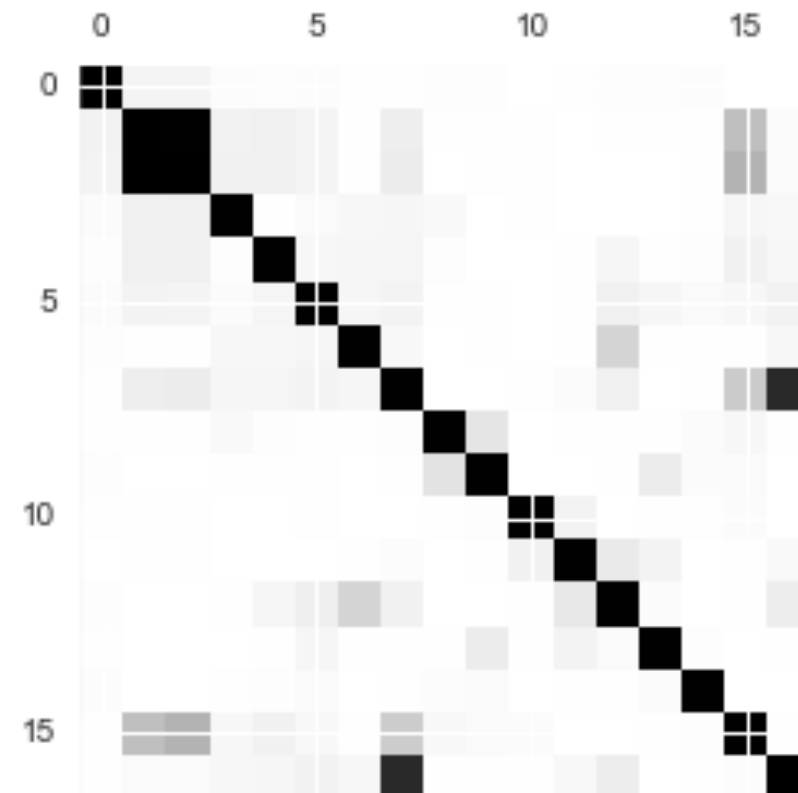
CORRELATION

We will now look at the correlation between different parameters in the dataset.

We can see in the attached figure, the correlation matrix for all the parameters in the dataset.

This shows the (by definition) correlations between arrival and departure as well as price per night and total price.

We can now go on to explore and find more meaningful correlations in the data.



CORRELATION WITH PROMO TO BOOKING

- Question:
- Is there a correlation between search results having a promotional offer and booking being made?
- Answer:
- We have shown to a high probability (better than 95%) that a promotional code increases the chance of a hotel being booked.

This information can be used to suggest promotional offers to hotels by showing them how much it increases their likelihood of being booked.

CORRELATION BETWEEN PRICE AND BOOKING

- ▶ Question:
- ▶ Is there a correlation between the price of a hotel per night and whether it is booked?
- ▶ Answer:
- ▶ By categorising the price per night into bands. We have shown that hotels in the band (30.37, 84.73) are more likely to be booked.
- ▶ This can be used by hotels to adjust their pricing if they are looking to make more bookings.

CORRELATIONS WITH PRICE (PER NIGHT)

- **Question:**
- Is there a correlation between the price per night and the hotel features?
- **Answer:**
- Our findings show that any correlation between hotel features and price is weak.
- Therefore hotels should not be relying on these features as a predictor of price.
- More features would be needed in order to begin predicting price.

CORRELATIONS WITH CUSTOMERS

- Question:
- Is there any correlation between the number of adults and children and the hotel features?
- Answer:
- By using a Chi-Squared test, we ascertained that there is no correlation between number of adults, or number of children with the hotel features.
- This suggests that none of the hotel features in the dataset are relatable to the child friendliness of the hotel or the hotel's appeal to adults.

SUMMARY

- ▶ We have found the common dates being booked and the most common length of stay
- ▶ We have found correlations between promotional codes, price categories and bookings
- ▶ We have found that there is not a correlation between hotel features and price
- ▶ We have found that there is not a correlation between hotel features and the number of adults and children.

AND FINALLY...

We made an attempt at a regression model to predict hotel prices based on features, however it was not a good fit.

This model would need access to more features and training time before it could be considered useful.

When a better model was built it could be used to predict hotel prices based on hotel features and from that determine if a hotel was under/over priced.

FURTHER ANALYSIS

- ▶ To perform more detailed analysis of the data it would be helpful to have access to further fields such as:
 - ▶ Number of rooms requested per search
 - ▶ Location of hotels
 - ▶ Grouping searches by user (taking into account returning users)