BROWN
DATA
SCIENCE
DATATHON

Whose Booking and Where are they Booking?

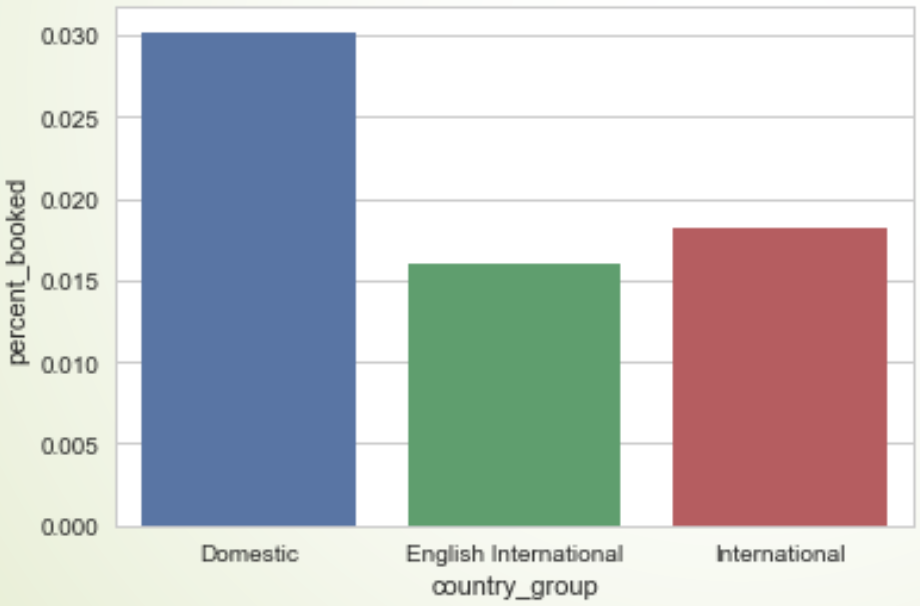
Victòria Gras Andreu, Peihong Jiang, Vikram Saraph, Ashley Weber

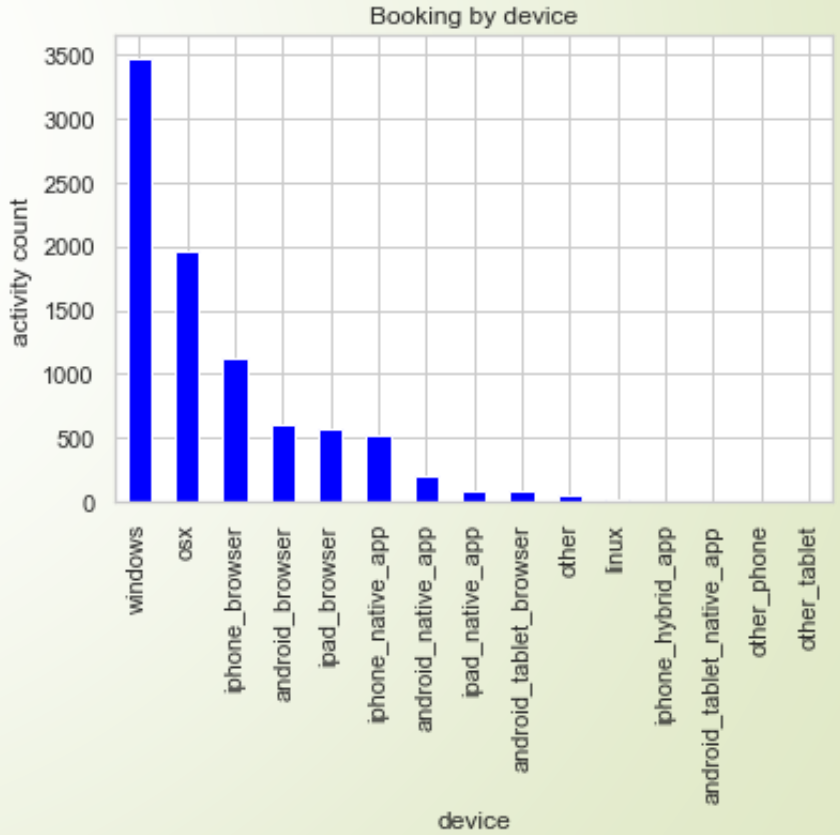
Goal 1: Predict if a user will make a booking

Can we predict it based on user activity and information only?

Data Exploration

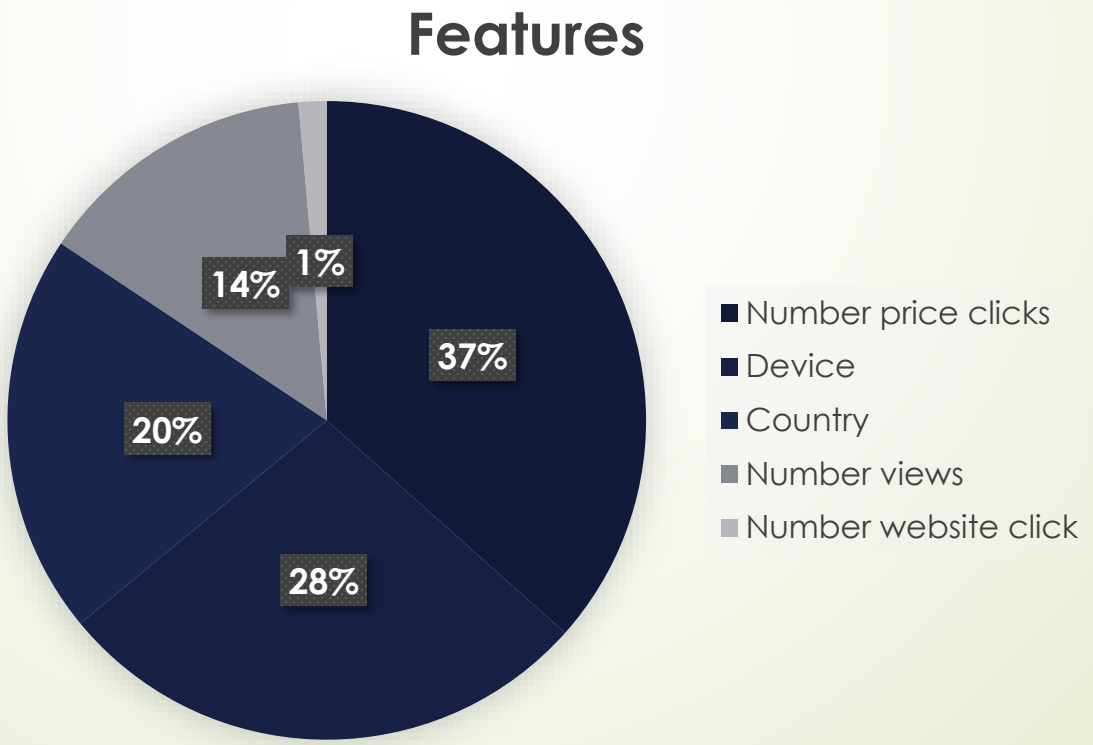
- Is there any relationship between bookings and countries?
- Are countries whose official language is not English more or less likely to book through TripAdvisor?
- Can we use the type of device to predict if a user will make a booking?





Model Training

- XGBoost Classifier
- Goal: Predict if a user has made a booking
- Accuracy: 97-98%
- Conclusion: We can predict if a user will make a booking based on user information only!

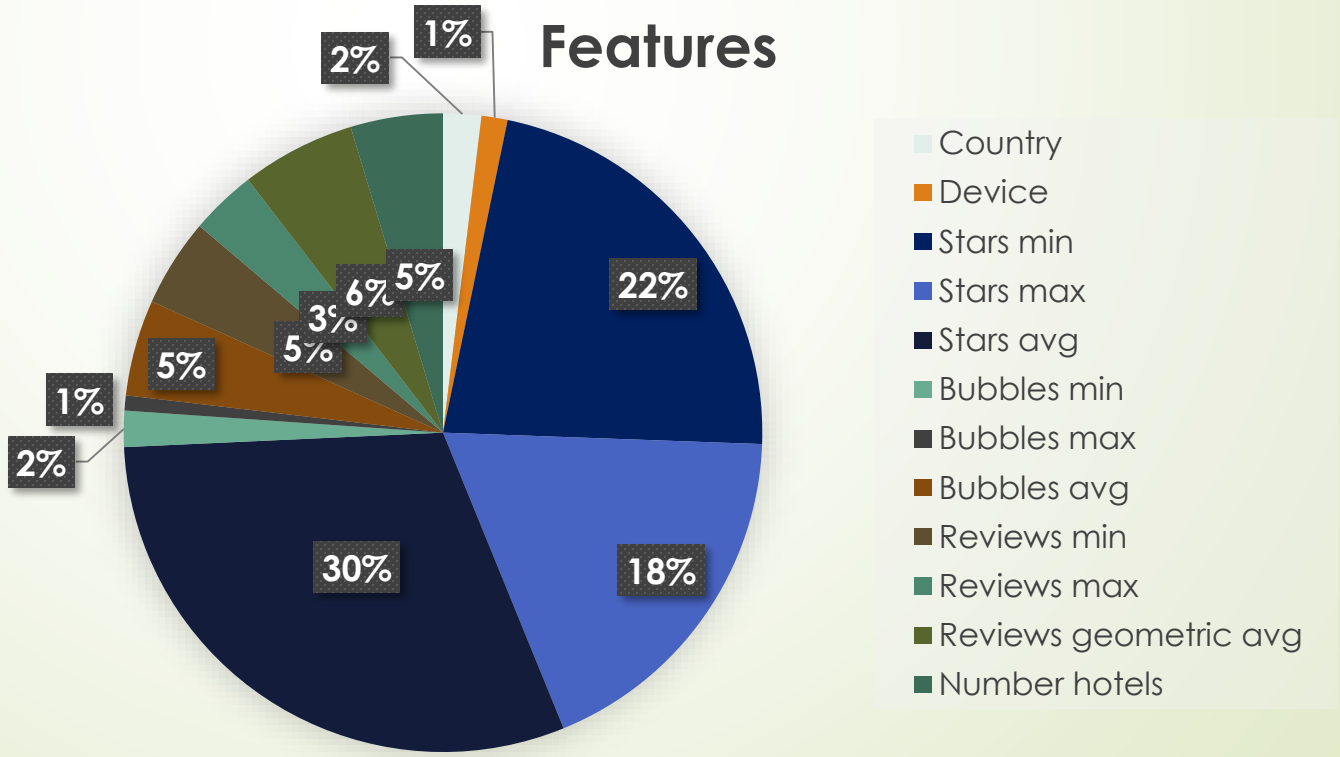


Goal 2: Predict characteristics of the hotel a user booked

Knowing a user has booked a room, can we predict the location and rating of that hotel?

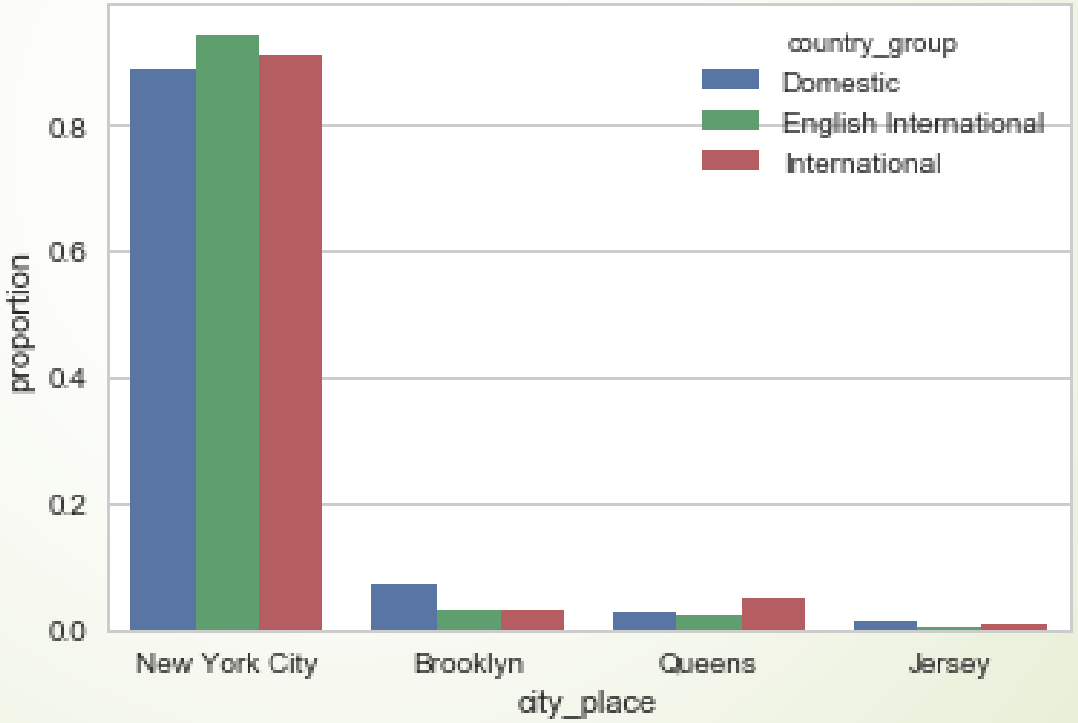
Star Ratings Predictions

- XGBoost Classifier
- Goal: Predict the star rating of a hotel a user has booked
- Accuracy: 79%
  - Observation: Allowing the prediction to be correct within 0.5 we get an accuracy of 93%!
- Conclusion: We can predict the star ratings of a hotel!



Prediction of location

- Do people from the US, countries where English is an official language, and countries where the English is not an official language book hotels in different locations?

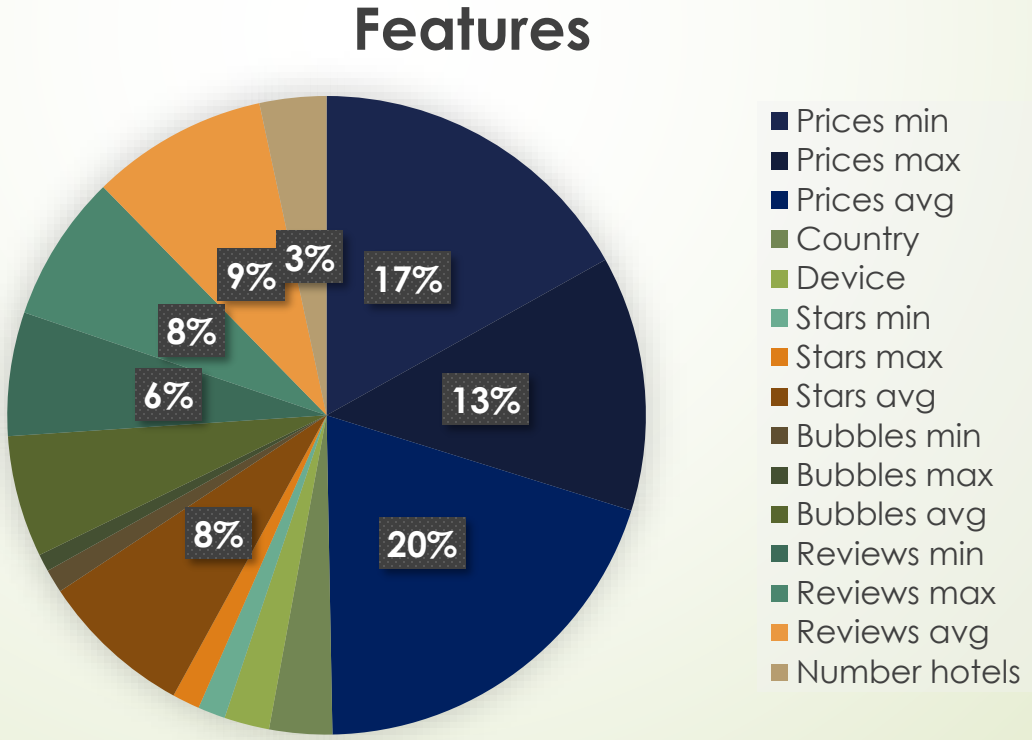


We need more data!

- Doesn't give a lot of insight on where people book within New York City!
  - Map of the most popular places in New York City
  - https://sites.google.com/brown.edu/demodemodemo/
- What about price versus location?
  - Assumption: there is a relationship between the price of the hotel and rent in each zip code.
  - High rent ➔ Price of hotel higher ➔ Most popular neighborhoods
- New data: Zip codes of the hotels (TripAdvisor website) and rent prices (Zillow).
  - Issues: We couldn't find the Zip code of all the hotels or rent prices for some Zip codes.
  - We had to drop some data! We used 62% of the data.

Zip Code Prediction

- XGBoost Classifier
- Accuracy: 71%
- Conclusion: Seems like we can predict locations but there's room for improvement!
- Prices are the most important features (blue), with more data we could do better!



Other projects

- With more data:
  - Using the user history for each day, predict what hotel they'll view next to ultimately show the hotel they'll likely to book (RNN)
    - Only half of the users viewed and booked on different days.
  - Use the click stream to predict what hotel they're more likely to view/book (RNN)
  - Given a hotel, predict how likely/how many times is booked using all the features available of the hotel with zipcode information, pricing of the zipcode area (XGBoost)

Thank you for your attention!