Assumptions:

- Data with the scraped_date > date is excluded from analysis

- Units with completely all unavailable dates are unlisted

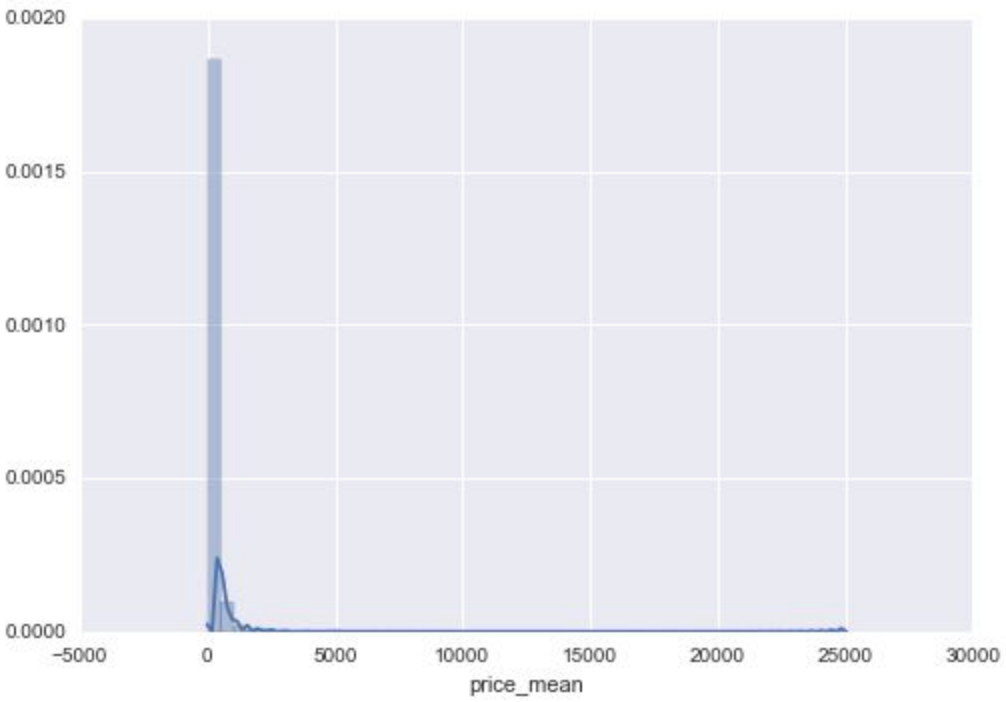
1. **What data would you exclude from analysis for being unreliable or potentially a block instead of an actual booking?**

    a. **Dates Far Away:**

        i. I would exclude dates that are far from the date of scraping as the dates close to the date of scraping would be most likely bookings. Available = 0 for four weekends out in advance or within two weeks of the date of scraping would most likely be actual bookings. This works the opposite way as well. If there exist days far out over a month from the day of the scraping that never become available, then it's likely that those days are blocked.

    b. **Ridiculous Pricing:**

        i. Excluding dates where the pricing is far above and is an obvious outlier. A way to determine this is to plot the distribution for average price of a unit segmented by number of bedrooms.



There's a bump at 25K that needs to be cleaned up

    c. **Looking at Reviews:**

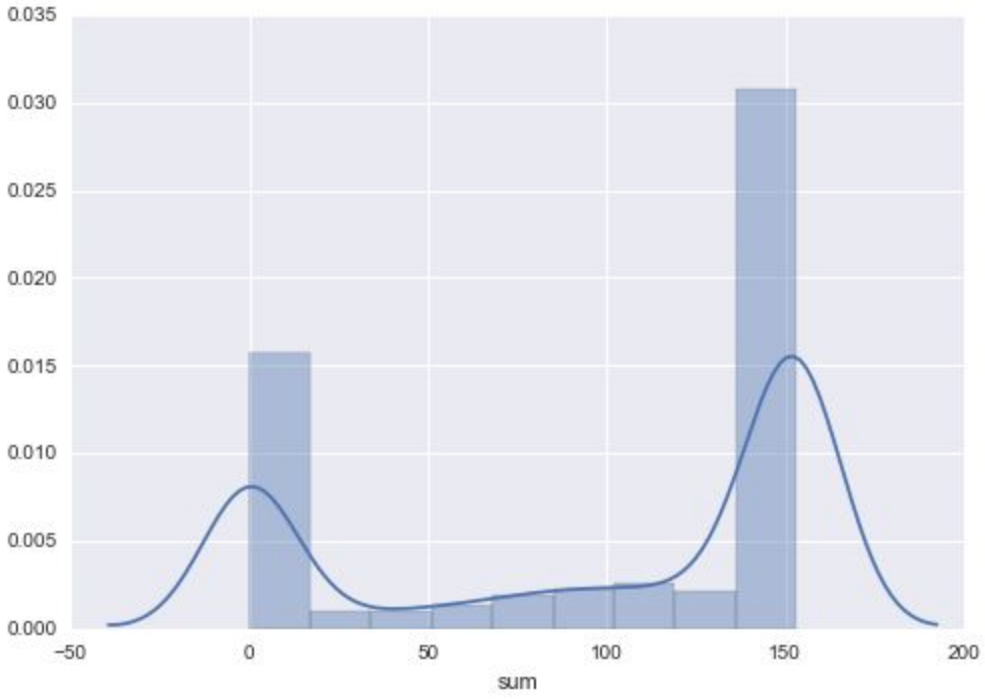
        i. If reviews were available, then looking at places with a high number of reviews and on the higher end of the distribution compared to other similar units would then conclude that all of the dates unavailable are

booked rather than blocked. Same could go for the other way around (many dates unavailable, no reviews, all probably blocked).

2. **What is a good approach to estimate occupancy and revenue per unit?**

Estimation of occupancy and revenue would require cleaning up and breaking up the dataset and calculating multiple parts:

- **Dates first available then turned unavailable at another scraped date:**
  - These dates would be guaranteed occupancy.
  - Revenue would be these dates multiplied by price listed.
- **Filtering out**
  - **Units with outlier pricing or large bedroom units**
    - Units with outlier pricing or large bedroom units will both fall under a case of estimated low occupancy given how sparse these units are ever booked. A 10 bedroom place with a capacity of 40 people does not frequently get booked.
  - **Units with dates completely unavailable for all dates:**
    - Filter out these units as they are probably unlisted as it's impossible to determine if they are insanely popular Airbnb's or completely unlisted without looking at reviews.



- Graph displays the bimodal distribution of number of total available days for a specific view of a couple of scraped days.

- **Dates unavailable for the entire scraped period and since the first scraped date:**

- Given that there doesn't exist other strong indicators within the listing attributes, we would have to use an estimated probability score for the expected value of these dates being available or not.
- Probability score factors:
    - *# of dates first available then turned unavailable/ # of dates first available*
        - Tells us how popular the unit listing is. If a large number of initially available days become booked, then there's a higher probability the existing unavailable dates are from bookings.
    - *Max number of days out the unit was unavailable*
        - Tells us the likelihood of being booked versus blocked. If a unit is a week out from scrape date and unavailable, it's likely booked. If it's three months from the scrape date, then likely blocked.
- Expected value is then a multiple of these two probability factors times the unit price.

3. **Which month appears to be more profitable? April or May?**
    - May. Running through calculations I got estimates of:
        - 2018-04 1.376324e+07
        - 2018-05 1.397840e+07
4. **How much more revenue do places with 3 bedrooms make vs. places with 2 bedrooms?**

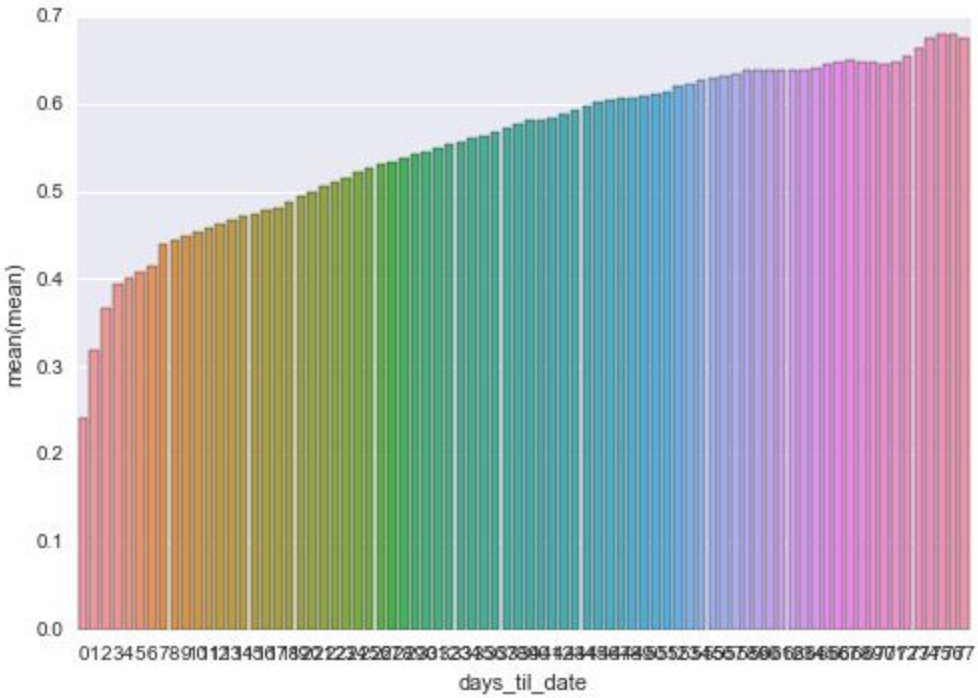
    Grouping together the stats:

Beds	Total Sum	Count	Revenue Per
2	6.839198e+06	1404	4871.223445
3	6.715126e+06	883	7604.899384

Looks like three bedroom units make around 2500-3000 dollars more on a per unit basis.

5. **What are any other interesting insights you may have found?**

Availability by days from booking



- Basically plan to book at least a week in advance or else the availability rapidly drops off. I will use this for my own benefit when planning vacations in the future.