AirBnB Research Proposal
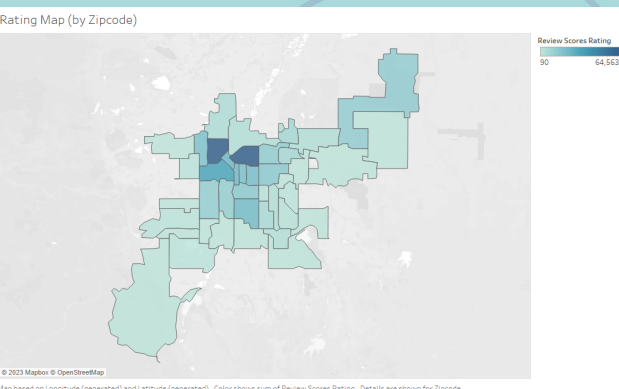
1

What are the top keywords mentioned in the textual reviews?
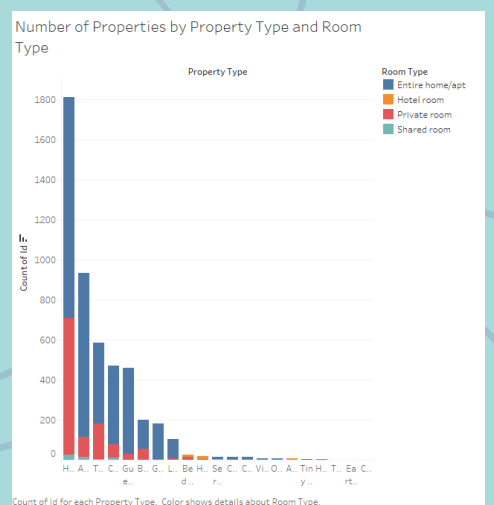
Rating Map (by Zipcode)

Review Scores Rating
30   64,563

Map based on Longitude (generated) and Latitude (generated). Color shows sum of Review Scores Rating. Details are shown for Zipcode.

2

How do the top keywords vary among property and room types?

3

Overall sentiment of all textual reviews? Sentiment by topic?

Number of Properties by Property Type and Room Type

Property Type

Room Type
Entire home/apt
Hotel room
Private room
Shared room

H. A. T. C. Su B. G. L. Be H. Se C. C. Vi O. A. Tin H. T. Ea C.

Count of Id for each Property Type. Color shows details about Room Type.

Room Type

Shared room
1.29%   Hotel room
              0.80%
                         Private room
                         23.51%

Entire home/apt
74.39%

4

How does sentiment vary across different neighborhoods?
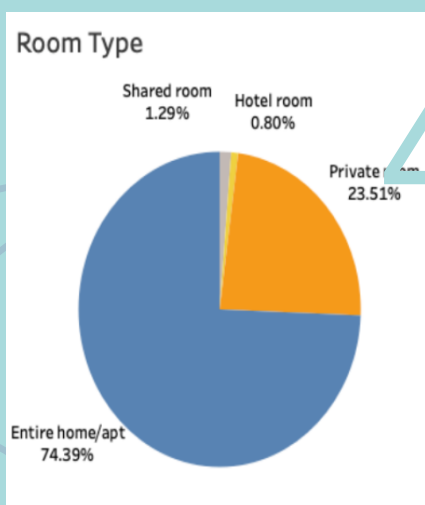
5

What topics are mentioned, and what keywords make up these topics?

6

Can we accurately use textual reviews to predict the price per day?

Variables

Variable - Description - Type
id - ID of the listing - Continuous, integer
listing_url - Url of the listing - Nominal, text string
name
Name - of the renting house/room -Nominal, text string
summary - Summary of the place - Nominal, text string
experiences_offered *(could be removed) - Whether the listing also offers other traveling experiences - Nominal, text string
*(should be binary)
interaction - The host's desired way of interaction - Nominal, text string
house_rules - House rules - Nominal, text string
host_id - The ID number of the host - Continuous, integer
host_name - Name of the host - Nominal, text string
host_since - Since when the host has been hosting - Continuous, integer *(should be date)
host_response_time - Host response time - Nominal, text string
host_is_super_host - Whether or not the host is a super host - Nominal, text string *(should be binary)
host_identity_verified - Whether or not the identity of the host is verified - Nominal, text string *(should be binary)
neighborhood - The neighborhood the house is at - Nominal, text string *(should be categorical)
zipcode - The zip code of the listing - Continuous, integer
property_type - The type of the property - Nominal, text string *(should be categorical)
room_type - The type of the room - Nominal, text string *(should be categorical)
accomodates - Number of accommodates - Continuous, integer
bathrooms - Number of bathrooms - Continuous, numeric
bedrooms - Number of bedrooms - Continuous, integer
beds - Number of beds - Continuous, integer
bed_type - Type of the bed - Nominal, text string
amenities - Available amenities - Nominal, text string
square_feet - Total size of the listing in square feet - Continuous, integer
price - The listing price per night - Continuous, decimal
weekly_price - The listing price per week - Nominal, text string *(should be continuous, integer/decimal)
monthly_price - The listing price per month - Nominal, text string *(should be continuous, integer/decimal)
cleaning_fee - The fee for cleaning - Nominal, text string *(should be continuous, integer/decimal)
minimum_nights - Number of minimum nights for each reservation - Continuous, integer
availability_365 - Number of available days for the year - Continuous, integer
number_of _reviews - Number of reviews - Continuous, integer
first_review - First time the listing was reviewed - Continuous, integer *(should be date)
last_review - Last time the listing was reviewed - Continuous, integer *(should be date)
review_scores_rating - Overall rating of the listing - Continuous, integer
review_scores_accuracy - Average rating on accuracy - Continuous, integer
review_scores_cleanliness - Average rating on cleanliness - Continuous, integer
review_scores_checkin - Average rating on check-in - Continuous, integer
review_scores_communication - Average rating on communication - Continuous, integer
review_scores_location - Average rating on location - Continuous, integer
review_scores_value - Average rating on value - Continuous, integer
instant_bookable - Whether or not the listing supports instant booking - Nominal, text string *(should be binary)
cancellation_policy - The cancellation policy of the listing - Nominal, text string
calculated_host_listings_count - The number of house/rooms owned by the listing's host - Continuous, integer
reviews_per_month - Average number of reviews each month - Continuous, numeric