Data Analytical Application Development

Airbnb Assignment

1. Read the Texas Airbnb File into a pandas DataFrame and then use the first column as index.

2. Drop records that have a missing value in any of average\_rate\_per\_night, bedrooms\_count, date\_of\_listing, latitude, and longitude.

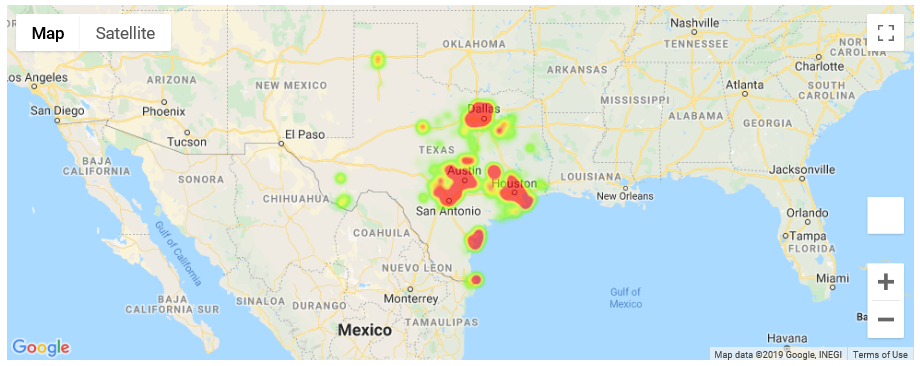
3. There is any negative number in average rate column? If yes, remove the related rows.

4. What is the average price of 2-bedroom listing? Write statements to calculate it.

5. In the ORIGINAL DataFrame, add TX to the end of all cities in the city column. For example, if the city is Houston, the new content is Houston, TX. Display the first 5 rows of the DataFrame.

6. Display the average nightly price of 2-bedroom grouped by each city.

7. Display a heatmap on Google map for all listings. What are the five popular areas for Airbnb listings in Texas?



8. Ask the user to enter a city and the number of bedrooms. Your program will display an average nightly rate based on the user’s input.





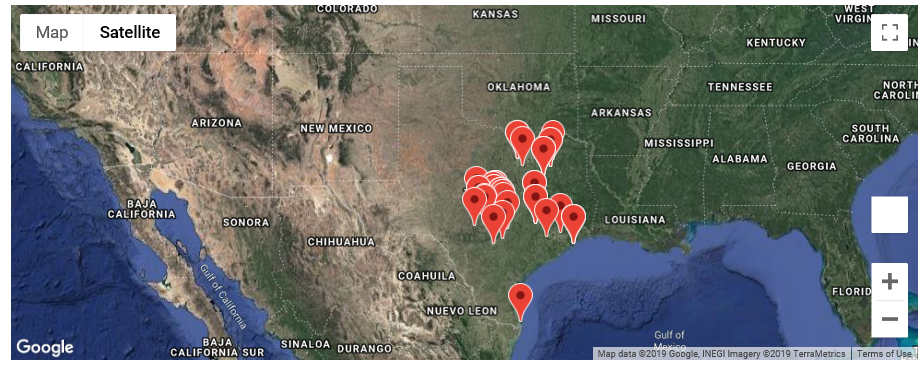
9. Write statements to display the numbers of listings of each bedroom type, from the highest to the lowest. Which room type has the largest number of listings?

10. Write statements to display the average nightly rate for each city where it has at least 100 listings.

11. Use the output in the last question to draw a price heatmap on Google map. In a price heatmap, a hotter area means the average nightly rate is higher in the area. Where is the area that has the highest average nightly rate with at least 100 listings?

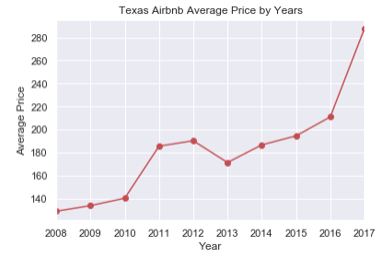


12. On a Google map, please mark the locations for the listings which has more than or equal to 8 bed rooms. Add a text notation with the number of bedrooms and price to each marker.

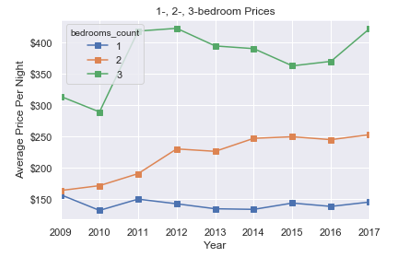


13. Group the listings by city where each city has at least 50 1-bedroom listings. Display the top 10 cities where they have the largest increase of nightly price between 2016 and 2017. Display the data with a descending order on the difference, along with their 2016 and 2017 average nightly rate of 1-bedroom listings.

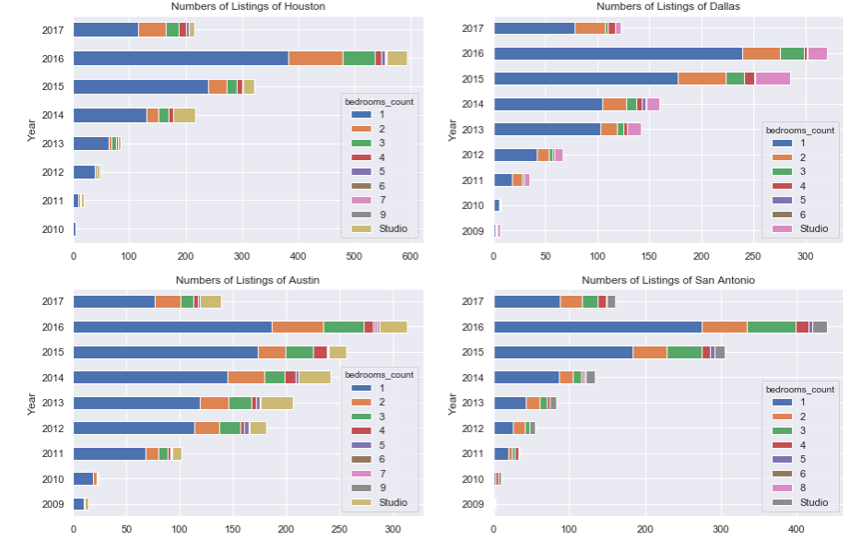
14. Draw a line chart for displaying the average nightly price (of all types) by years. The line chart needs to have a red color line, a title, and x, y labels.



15. Draw a line chart for displaying the average nightly prices of 1-bedroom, 2-bedroom and 3-bedroom by years. Since there is no data for year 2008 for 2-bedroom and 3-bedroom, the years start from 2009.

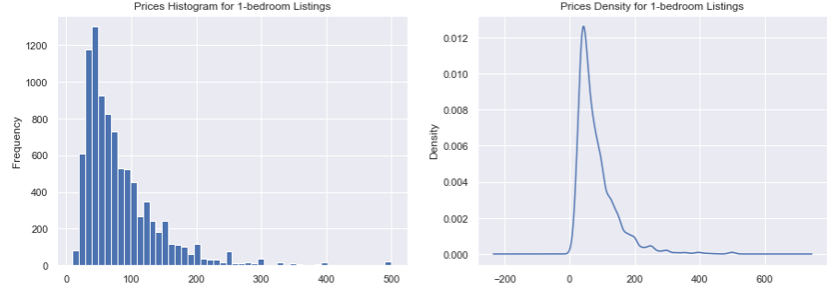


16. Create a figure which contains four subplots. Each subplot contains a stacked columns chart of total listings by years. The four subplot charts are for data of Houston, Dallas, Austin and San Antonio.

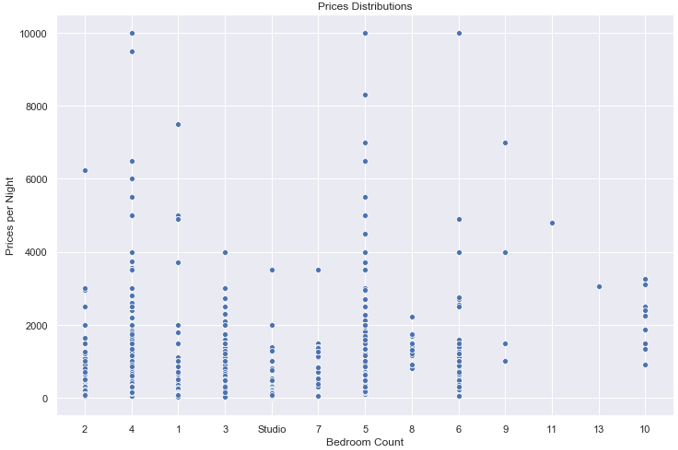


What you can tell (3 of your own opinions) from these four charts? Please ignore that the numbers of listing in 2017 that obviously smaller than the ones in 2016. I believe that the reason is because data of 2017 was not completely collected when the data set was created.

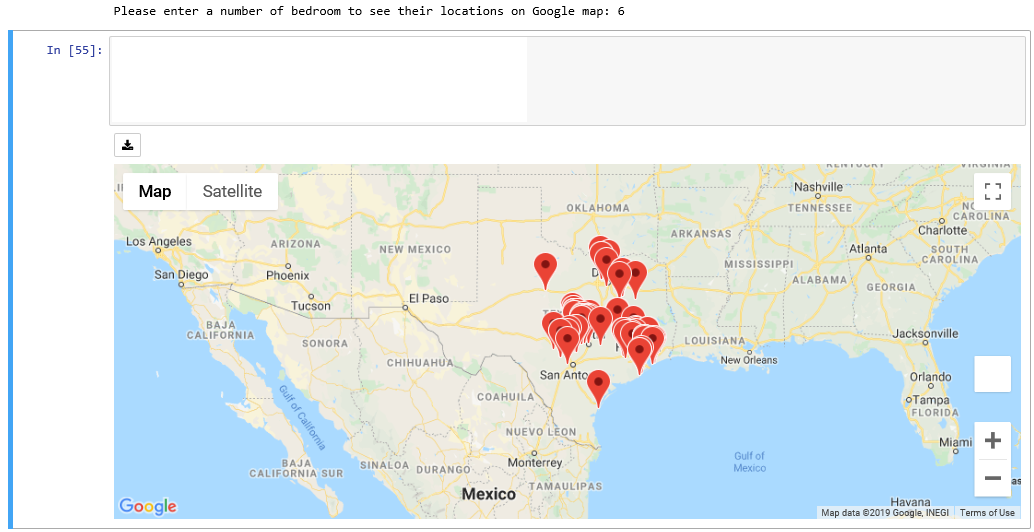
17. Create a figure with two subplots. The first subplot contains a histogram chart of 1-bedroom listings where the nightly rates is between $0 and $500. Use 50 bins for the histogram chart. The second subplot is a probability density chart of the same data.



18. Please use a scatter plot to display the price distribution of all listings based on different bedroom numbers.



19. Ask the user to enter a number of bedrooms. Display a Google map of all of the listings with that type of bedrooms marked. Use a number of at least 5 to test your program because small-house listings are too many.



20. Enjoy your completion of this project. I hope you are more confident now!