horizontal line

**Sam Gentry Data Analysis Portfolio Project**

San Diego Airbnb Data Exploration

**14th May 2023**

# OVERVIEW

The Dataset: Found on <http://insideairbnb.com/>

Tableau dashboard (Shows usage of Joins, graphs and viz construction) - <https://public.tableau.com/app/profile/samuel.gentry/viz/AirbnbSanDiegoProject/Dashboard1#1>

GSheets data and charts (Shows pivot, calculated fields and charting) - [AirBnB San Diego Listings](https://docs.google.com/spreadsheets/d/1cE62n1votHz5jqvUN1CsW5enntfZe2OQ7auYFNvr7K4/edit?usp=sharing)

I downloaded this data from the website above and focused on two data sets; Listings and Reviews. In order to gain some knowledge about popular listings, investment opportunities, frequency of reviews and other market knowledge.

The Schema:

The listings table contains unique Id, price per night, description, room type, number of reviews, ave per month reviews and neighborhood information. There are two other tables I reviewed and drew on for information; detailed listings with much more data on reviews and hosts along with the neighborhood geo table which gave me location of the neighborhoods for mapping.

# GOALS

1. Determine the area that has the most rentals and is the most successful financially.
2. Determine what type of rental (full house, room, full year avail or seasonal) is most successful.

# Concerns/Constraints

The data used to determine the financial success is not direct but rather an estimated calculation using number of reviews \* minimum length of stay \* price per night. So financially our analysis could have some root concerns given the constraints of the data we have available.

# Data Analysis Process

## Ask

During this phase I explored the data and began to ask myself some questions in order to derive an overall strong business question to solve. What neighborhood is most popular? How can we quantify popularity? (ratings, open dates through the year, pricing). Do factors like 1 room vs. whole house make a difference in that ‘popularity’?

## Prepare

At this phase I made a copy of the original table and started my own with deleting a few columns I intended to not use in my analysis; one was fully blank about neighborhood type and another was about licensing which I don’t intend to focus on in my analysis. I went through the other columns sorting and filtering to check the validity and cleanliness of the data. During this exploration, I decided to delete one observation due to the pricing for that listing being $100k for one room while the next closest listing’s were $17k. This seemed to be an error or a joke/price change from the host in my opinion. I also reviewed the averages, min, max of all the numerical data to determine the shape of my data set.

## **Process**

After the removal of unnecessary columns and apparent mistaken observation/listing price. Here I began to gain some insight to the data available here and how to create a popularity or financial success equation. I decided to take the ‘reviews per month’ and multiply that by 12 months to create an average of reviews per year. That value becomes the assumed amount of rentals for the year. I will take that value and multiply by the minimum day stay and then that number by the price per night to determine gross revenue for the year.

**Analyze**

With my question in mind of the most popular properties/neighborhoods for Airbnb in San Diego, I constructed a pivot table containing some of the table data aligned with the various neighborhoods. I was able to compile this pivot and factor in the calculations I indicated in the Process phase.

**Share**

Once I had my pivot table completed I was able to create some charts to display the neighborhood revenue rankings in G-Sheets. I took this data along with the original data and some geographic data into Tableau to share some mapping and graphing visualizations on a compiled dashboard.

**Act**

Having reviewed the data and comparing the number of listings, price per night and yearly gross revenue I see some correlations. Firstly the most popular area concerning those three things is by the ocean in the central suburban area of San Diego. Second most popular area is the downtown entertainment area near the MLB ballpark. My personal thoughts would be to invest in a neighboring area to the most popular because the cost to purchase should be less while you can still reap the rewards of this popularity. Neighborhoods I’d choose from would be Ocean Beach, Wooded area and East Village.