

# Airbnb data analysis

Identifying most common amenities that give Airbnb hosts a competitive advantage over other hosts in Gauteng & Cape Town, South Africa

~Henrick Aduda



The Data Incubator

**Inside Airbnb**  
Adding data to the debate

# Motivation

- Airbnb hosts took a massive hit in 2020
- Airbnb revenues decreased by 48.6% (\$35 billion in Jan to \$18 billion in May)
- Opportunity: hosts in rural areas of the U.S. earned over USD 200 million in June 2020
- Motivation : find approaches/means for airbnb hosts to get more clients.



# Data source description

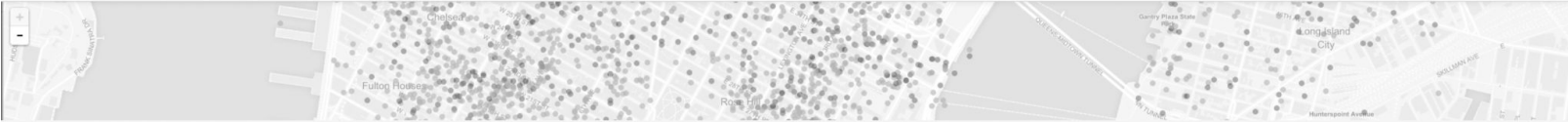
- Original dataset came from Inside Airbnb web database
- Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world.

<http://insideairbnb.com/get-the-data.html>

**Inside Airbnb**  
Adding data to the debate


AboutBehindGet the Data

Like 5.4KShareTweet



## Get the Data

The data behind the Inside Airbnb site is sourced from publicly available information from the Airbnb site. The data has been analyzed, cleansed and aggregated where appropriate to facilitate public discussion. See more [disclaimers here](#), and a data dictionary [here](#).



If you would like to do further analysis or produce alternate visualisations of the data, it is available under a [Creative Commons CC0 1.0 Universal \(CC0 1.0\) "Public Domain Dedication"](#) license.

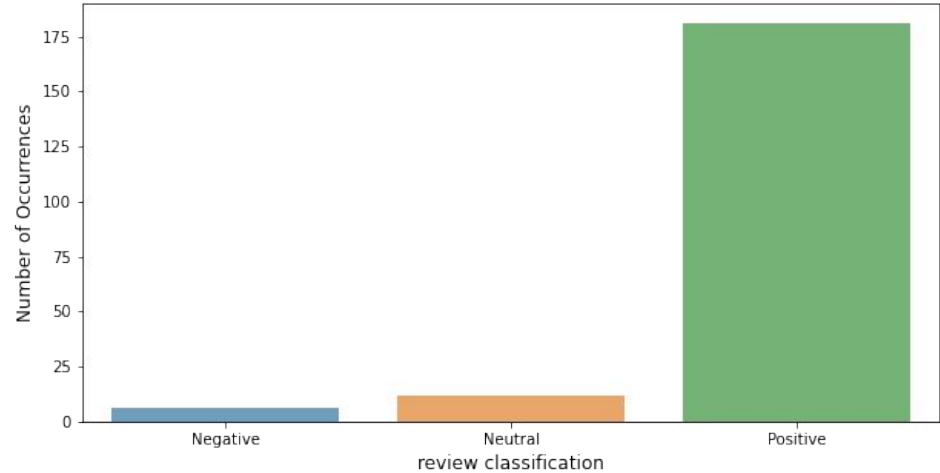
If you have any questions, or would like to request data you don't see here, please contact [data@insideairbnb.com](mailto:data@insideairbnb.com). Please let us know who you are, your interest in the data and Airbnb. We prioritise requests based on their alignment with the project's mission - to provide free data that quantifies the impact of short-term rentals on housing and residential communities; and also provides a platform to support advocacy for appropriate and effective policies to protect our cities from the impacts of short-term rentals.

# Analysis

## Deliverables

1. Identify the most common amenities affecting review ratings - correlation analysis
  2. Perform a sentiment analysis - monkeylearn web tool - aim for ML model accuracy >80%
  3. Identify the most common amenities affecting review comments - Yet to figure out plan
- **Opportunity:** Scale it up to look at world wide data. Analysis per continent

## Airbnb sentiment analysis reviews



200 out of 305085 review entries

## Confidence mean scores

Positive : 0.97

Negative: 0.83

Neutral: 0.69

All entries: 0.97