

# User Guide for Business Mafia's RShiny Application

Too High, Too Low, or Just Right? ☰

Overview

- Background
- Sandbox (Data Sources)
- Point Pattern Analysis
- Geographically Weighted Regression

Hi there.

**Here's 2 questions. How can an Airbnb host have a more perfect knowledge of the market around her/his listing(s), regardless if it's existing or new? And what should be a good price of leasing one for a night?**

A majority of Airbnb hosts are individuals who list apartment(s) as side business, but they generally lack the knowledge to build optimal pricing models for themselves. And while Airbnb has made its data publicly available, there is still a general lack of analytical platforms to guide hosts in answering the above questions.

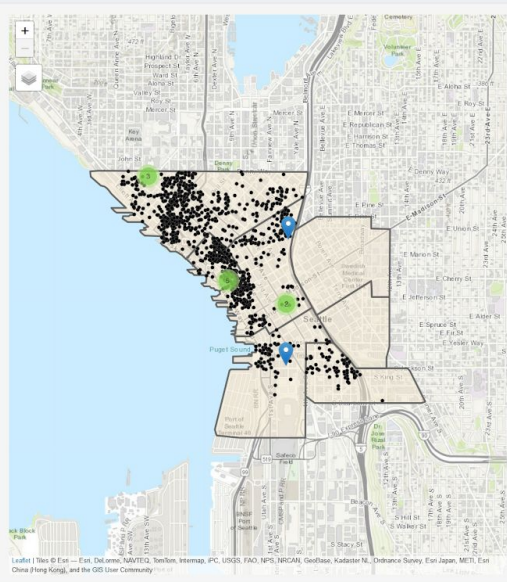
Our team feels that any hosts can benefit from the distillation of these big data into valuable insights. Democratizing such data through web based applications such as ours could trigger better decision making for hosts all around, which in turn makes the market more competitive; spilling over its benefits to consumers like you and us.

Hence, our model aims to stand in this gap with the following services:

1. A picturesque and interactive map application, powered by Leaflet, that invites users to derive new insights on the neighbourhood(s) they've chosen, and;
2. A Geographically Weighted Regression (GWR) that helps hosts more confidently price their listings at optimal rates, rather than arbitrarily.

**Data Sources used:**

1. Airbnb Listings  
A .csv file originally containing information on all Airbnb listings found within Seattle, last scrapped on 15 November 2018. With this file, we filtered for listings based only in Downtown Seattle
2. Common Place Names  
A shapefile featuring common place names and corresponding locations in Seattle.
3. City Clerk Neighbourhood  
A shapefile featuring 20 Large City Clerk neighborhood boundaries, along with their smaller neighborhood boundaries.
4. Zoning (Generalised)  
A shapefile featuring boundaries which are not labelled by neighbourhood names, but rather, urban planning descriptions. I.e. Downtown, Major Institutions, Manufacturing/Industrial, Multifamily, Neighbourhood/Commercial, Residential/Commercial and Single Family



## Main Page of Our App

Welcome to Business Mafia's RShiny Web Application user guide book! When you first launch our app, you should reach this page. This page contains a short write-up of our project, our objectives and wishes when designing this app. Our app serves not only current Airbnb hosts but also individuals who are looking to become Airbnb hosts themselves in the future. In the main page, the extreme left-most bar on the side contains the various main features of our app. It has: Background, Sandbox (where you can peek and look at the different data sets that we used) and two key geospatial analytical techniques that hosts/prospective hosts might be interested in. Also, on the right most of our the screen shows all Airbnb apartments listed in Downtown Seattle (represented by the small black dots) and 11 key locations within Downtown Seattle that we have identified. They are represented by the larger green circles and blue arrows.

Seattle Listings   Common Place Name   City Clerk Neighbourhood   Zoning Generalized

Source: insideairbnb.com

Show 10 entries

id	host_since	host_response_time	host_is_superhost	host_identity_verified	property_type	room_type	accommodates	bathrooms	bedrooms	bed_type	price	minimum_nights	maximum_nights	
20	19619	2010-01-27	within an hour	0	1	Condominium	Entire home/apt	4	1	1	Real Bed	145	2	365
26	24212	2010-01-27	within an hour	0	1	Condominium	Entire home/apt	6	1	2	Real Bed	157	1	365
40	59827	2010-01-27	within an hour	0	1	Apartment	Entire home/apt	3	1	1	Real Bed	96	2	365
76	202174	2009-11-28	within an hour	0	1	Townhouse	Entire home/apt	4	2	2	Real Bed	159	2	730
79	208356	2011-08-27	N/A	1	1	Loft	Private room	2	1.5	1	Real Bed	99	4	14
83	217142	2013-08-18	within a day	1	1	Condominium	Entire home/apt	7	2	2	Real Bed	200	2	365
87	224763	2013-08-18	within a day	1	1	Condominium	Entire home/apt	4	1	1	Real Bed	139	1	365
107	271509	2010-01-27	within an hour	0	1	Condominium	Entire home/apt	6	1	2	Real Bed	175	2	365
108	277416	2013-08-18	within a day	1	1	Condominium	Entire home/apt	4	1	1	Real Bed	139	2	365
116	286080	2010-01-27	within an hour	0	1	Condominium	Entire home/apt	6	2	2	Real Bed	144	2	365

Showing 1 to 10 of 1,148 entries

Previous 1 2 3 4 5 ... 115 Next

Seattle Listings   Common Place Name   City Clerk Neighbourhood   Zoning Generalized

### Project Sandbox

Within the project sandbox tab, you will find four sub-tabs showing the different datasets that we used for our project. They are: Seattle Listings, Common Place Name, City Clerk Neighbourhood and Zoning Generalised. Don't forget to scroll to the right of the page for some data sets have way more columns then others. Feel free to click on them to explore the data sets!

### Description

We believe that the best decisions we make are the ones most well-informed. And our team wants to help you get there.

This tab will show you your neighbours' prices and characteristics, as well as a statistically derived value of when clusters of listings would form. With this information, we believe you are empowered to differentiate yourself from competitors, or be as similar to them.

Go on, be empowered.

### 2nd Order Point Pattern Analysis

You chosen location lies in the neighbourhood of Pike-Market

The graph on the left is the statistical results of a 2nd Order Point Pattern Analysis performed in this neighbourhood.

Where the black line crosses above the grey horizontal bandwidth indicates the minimum distance at which clusters of airbnb listings would begin to form.

### Leaflet Map

### Where to investigate?

You may enter an existing listing address, or a new listing address you're considering.

Enter the address

Run

### Filters

Go on, try it. And discover some new things about your location.

Choose your buffer:

0 200 400

Colored\_by

Property Type

### Buffer Analysis

Minimum buffer radius: 4 m

Number of neighbouring listings within buffer: 157

Property Type	Count
Apartment	119
Boutique hotel	1
Campsite/RV	5
Condominium	7
Hostel	1
Loft	1
Serviced apartment	1
Townhouse	1

The histogram above describes the distribution of listing prices within your chosen buffer.

The bar chart above reflects the proportion of a variable that your chosen buffer has.

## Point Pattern Analysis

Under the Point Pattern Analysis tab, users can experience the first Geospatial technique we introduced to our app. Based on radius around an address (also known as bandwidth in Geospatial terms), you can see the distribution of airbnb listings. You can change the bandwidth size to suit your liking, or to change the colouring on the map output. Also, the Second Order Point Pattern Analysis shows users that at which minimum radius does clustering patterns start to appear. However, the application is only limited to addresses within Downtown Seattle. Snippets of writeups by the beneath/side of the graphs show how the graphs can be interpreted!

### 2nd Order Point Pattern Analysis

You chosen location lies in the neighbourhood of Pike-Market

The graph on the left is the statistical results of a 2nd Order Point Pattern Analysis performed in this neighbourhood.

Where the black line crosses above the grey horizontal bandwidth indicates the minimum distance at which clusters of airbnb listings would begin to form.

### Buffer Analysis

Minimum buffer radius: 4 m

Number of neighbouring listings within buffer: 157

Property Type	Count
Apartment	119
Boutique hotel	1
Campsite/RV	5
Condominium	7
Hostel	1
Loft	1
Serviced apartment	1
Townhouse	1

The histogram above describes the distribution of listing prices within your chosen buffer.

The bar chart above reflects the proportion of a variable that your chosen buffer has.

Too High, Too Low, or Just Right? ☰

Overview

- Background
- Sandbox (Data Sources)
- Point Pattern Analysis
- Geographically Weighted Regression

Build your GWR

Select Variable to view its histogram plot:  
Price

Number of bins for histogram plot  
Slider: 1 to 28 (set to 4)

Select Bandwidth:  
Adaptive

Select Kernel:  
Boxcar

Choose your variables for GWR:

- No. of Guests/Living Accommodates
- Host's Identity Verified
- No. of Bathrooms
- Minimum No. of Nights
- Number of Reviews
- Amenities Index
- Walking Dist to Attractions Grp 1
- Walking Dist to Attractions Grp 2

Perform Regression

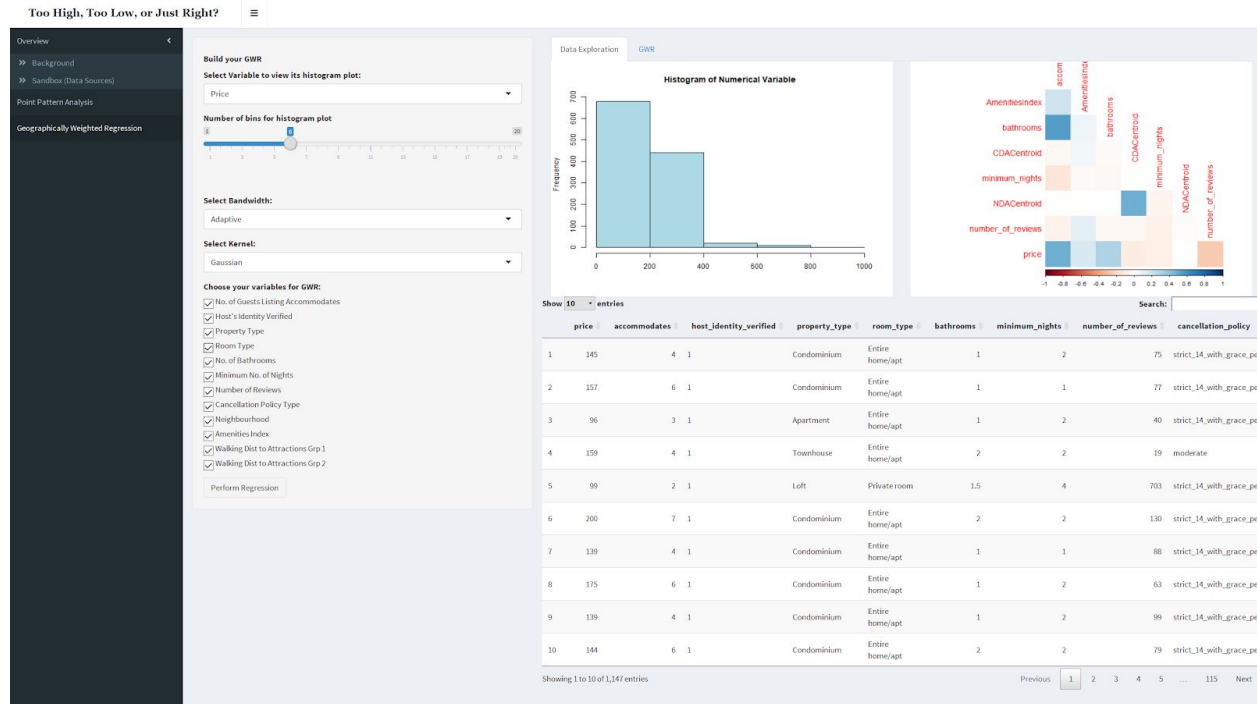
Data Exploration GWR

Histogram of Numerical Variable

Price Bin	Frequency
0 - 200	650
200 - 400	450
400 - 600	50
600 - 800	20
800 - 1000	10

## Introduction to Geographically Weighted Regression (GWR)

There are two sub tabs under the GWR model. We are currently under the 'Data Exploration' tab with nothing selected. First, you can select the variable(s) that you will like to see its histogram plots for. Also, the number of bins for the histogram plots can be chosen in the slider bar. Simply slide the bar to change the number of bins for that histogram plot. Something to note is that histogram plots are only available for numerical variables!



### Data Exploration tab with Corrplot feature

After looking at the individual histogram plot, another feature that we have introduced is the correlation plot function. The correlation plot provides you with a visual representation of the correlation value (r value) of the different variables chosen for GWR modelling. Checking for multi-collinearity between variables is an important step for regression modelling.

**Select Bandwidth:**

Adaptive ▼

**Select Kernel:**

Gaussian ▼

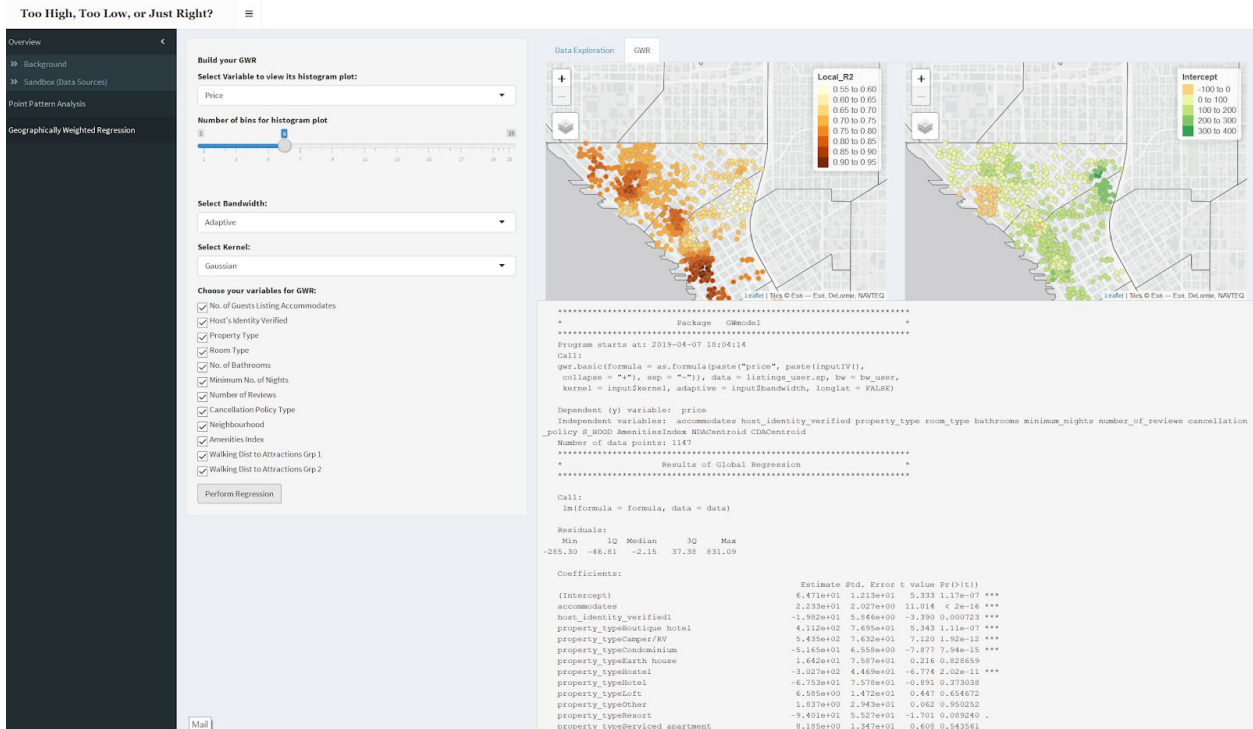
**Choose your variables for GWR:**

- No. of Guests Listing Accommodates
- Host's Identity Verified
- Property Type
- Room Type
- No. of Bathrooms
- Minimum No. of Nights
- Number of Reviews
- Cancellation Policy Type
- Neighbourhood
- Amenities Index
- Walking Dist to Attractions Grp 1
- Walking Dist to Attractions Grp 2

Perform Regression

### Calibration of Model Parameters for GWR modelling

The final feature in our application is the GWR. After exploring the different variables suitable for GWR, select the bandwidth type to use and the type of kernel function that you will like to use. Once ready, click on “**Perform Regression**”. The calculation takes approximately 30 to 60 seconds so do be patient! To view the results of GWR, click on the GWR sub tab.



## Geographically Weighted Regression model output

This tab shows the output and results of GWR model that we ran earlier. The graph on the left shows the local R2 value plot while the graph on the right the intercept value of individual AirBnbs independent of the explanatory variables. Lastly, at the bottom of the page, it shows the output of our the GWR model; both global and local regressions. Scroll down to have a better understanding of how spatial location plays a role in the regression modelling!

Happy surfing our application!