

Datasets Used

Seattle Airbnb: <https://www.kaggle.com/airbnb/seattle>

Seattle King County House Prices: <https://www.kaggle.com/harlfoxem/housesalesprediction>

Cleaned Dataset File Naming

Datasets Used File Name	Cleaned Dataset File Name
listings.csv	listings_cleaned.csv
reviews.csv	reviews.csv (No change)
kc_house_data.csv	kc_house_data_cleaned.csv

- R-script and Python script still uses the original file and has code to clean and visualize the data

Seattle Airbnb

listings.csv Data Dictionary

Variables/Factors	Description
id	Unique listing ID
listing_url	URL of the listing
scrape_id	Unique ID generated by Airbnb for their internal use
last_scraped	The date which the file was last updated
name	Title of the listing
summary	Short summary of description field to describe their house
space	Short description on the space around housing area
description	Full description on housing and surroundings
experiences_offered	Special experience offered to guests
neighborhood_overview	Description on neighbourhood features
notes	Extra information that hosts want to highlight

transit	Nearest mode of transport that guests can use
thumbnail_url	URL of listing's thumbnail
medium_url	URL of listing's medium
picture_url	URL of listing's picture
xl_picture_url	URL of listing's XL picture
host_id	Unique host ID
host_url	Link to host's profile
host_name	Name of host
host_since	Start date of when host started hosting
host_location	Location of where the host stays
host_about	Short description about host
host_response_time	Average time taken for host to reply queries
host_response_rate	Host percentage of queries responded to
host_acceptance_rate	Host Acceptance rate of bookings
host_is_superhost	Is the host a superhost?
host_thumbnail_url	URL for host's thumbnail
host_picture_url	URL for host's picture
host_neighbourhood	Neighbourhood of host
host_listings_count	Total count of host's listings for this property
host_total_listings_count	Overall total count of host's listings
host_verifications	Platforms that have verified the host's identity
host_has_profile_pic	Does the host have a profile picture?
host_identity_verified	Has the host verified their identities with Airbnb?
street	Street of listing

neighbourhood	Neighbourhood of listing
neighbourhood_cleansed	Cleaned user input of neighbourhood
neighbourhood_group_cleansed	General neighbourhood of the listing
city	City of listing
state	State of listing
zipcode	Zip Code of listing
market	State of the market
smart_location	City and State of listing
country_code	Country code of the listing
country	Country of the listing
latitude	Latitude of location
longitude	Longitude of location
is_location_exact	Boolean type on whether the location provided is the exact location
property_type	Type of the property
room_type	Type of room listed out
accommodates	Maximum number of people allowed
bathrooms	Number of bathrooms
bedrooms	Number of bedrooms
beds	Number of beds
bed_type	Type of beds provided
amenities	Amenities provided
square_feet	Size of the house in square feet
price	Daily price for minimum number of nights
weekly_price	Price for booking the house for a week

monthly_price	Price for booking the house for a month
security_deposit	Deposit amount for each booking
cleaning_fee	Fee for cleaning services
guests_included	Number of guests allowed
extra_people	Extra charges for each additional person
minimum_nights	Minimum number of nights per booking
maximum_nights	Maximum number of nights per booking
calendar_updated	When was the listing's calendar last updated
has_availability	Does the listing have available dates for booking?
availability_30	How many days out of the next 30 days are available?
availability_60	How many days out of the next 60 days are available?
availability_90	How many days out of the next 90 days are available?
availability_365	How many days out of the next 365 days are available?
calendar_last_scraped	The date which the calendar was last updated
number_of_reviews	Total number of reviews
first_review	Date of the first review
last_review	Date of the last review
review_scores_rating	Average score calculated based on the rating given by customers
review_scores_accuracy	Average score calculated based on the accuracy of the advertised listing
review_scores_cleanliness	Average score calculated based on the cleanliness of the house
review_scores_checkin	Average score calculated based on the check-in process

review_scores_communication	Average score calculated based on the communication experience
review_scores_location	Average score calculated based on the location of listing
review_scores_value	Average score calculated based on the value of the listing
requires_license	Is license required?
license	If licenses are required, which specific ones are they?
jurisdiction_names	The State of jurisdiction the rooms are under
instant_bookable	Can the rooms be booked on the day itself?
cancellation_policy	Category of cancellation policies: ['Flexible', 'Moderate', 'Strict']
require_guest_profile_picture	Does the host require the guest to have a profile picture?
require_guest_phone_verification	Does the host require the guest to have a verified phone number?
calculated_host_listings_count	How many listings does the host have
reviews_per_month	Number of reviews per month

listings_cleaned.csv Data Dictionary

Variables/Factors	Description
id	Unique listing ID
host_is_superhost	Is the host a superhost? [True, False]
neighbourhood_group_cleaned	General neighbourhood of the listing [Other neighborhoods, Capitol Hill, Downtown, Central Area, Queen Anne, Ballard, West Seattle, Rainier Valley, University District, Beacon Hill, Cascade, Northgate, Delridge, Lake City, Magnolia, Seward Park, Interbay]
latitude	Latitude of location

longitude	Longitude of location
room_type	Type of room listed out [Entire home/apt, Private Room, Shared Room]
accommodates	Maximum number of people allowed
bathrooms	Number of bathrooms
bedrooms	Number of bedrooms
beds	Number of beds
price	Daily price for minimum number of nights
security_deposit	Deposit amount for each booking
cleaning_fee	Fee for cleaning services
guests_included	Number of guests allowed
availability_365	How many days out of the next 365 days are available?
number_of_reviews	Total number of reviews
review_scores_rating	Average score calculated based on the rating given by customers
instant_bookable	Can the rooms be booked on the day itself?
Coded Variable/Factors	Description
property_type	Group type of the property into 3 property types [Houses, Apartments, Others]
sentiment	Average sentiment score of all reviews for the listing from reviews.csv
cluster	Cluster of the listing derived from KMeans Clustering with k = 3 on accommodates, bathrooms, bedrooms, beds, security_deposity, cleaning_fee and guests_included

reviews.csv Data Dictionary

Variables/Factors	Description
listing_id	Unique listing ID
id	Unique ID of review
date	Date of review
reviewer_id	Unique ID of reviewer
reviewer_name	Name of reviewer
comments	Reviews from the customer

Seattle King County House Prices

kc_house_data.csv Data Dictionary

Variables/Factors	Description
id	Unique listing ID
date	Date of listing
price	Selling price of house
bedrooms	Number of bedrooms
bathrooms	Number of bathrooms
sqft_living	Square foot of living room
sqft_lot	Square foot of plot of land
floors	Number of floors
waterfront	Whether the apartment has a waterfront view or not (1 for yes)
view	Score based on the view available from the house
condition	Score based on the current condition of the house
grade	Score based on the grade of the house
sqft_above	Square foot of real estate above ground level
sqft_basement	Square foot of the basement

yr_built	Year the house was built
yr_renovated	Year of the most recent renovation
zipcode	Zipcode of house
lat	Latitude of house
long	Longitude of house
sqft_living15	Adjusted square foot of living (Assumed)
sqft_lot15	Adjusted square foot of lot (Assumed)

kc_house_data_cleaned.csv Data Dictionary

Variables/Factors	Description
id	Unique listing ID
price	Selling price of house
bedrooms	Number of bedrooms
bathrooms	Number of bathrooms
sqft_living	Square foot of living room
sqft_lot	Square foot of plot of land
floors	Number of floors
waterfront	Whether the apartment has a waterfront view or not (1 for yes)
view	Score based on the view available from the house
condition	Score based on the current condition of the house
grade	Score based on the grade of the house
sqft_above	Square foot of real estate above ground level
sqft_basement	Square foot of the basement
yr_built	Year the house was built
yr_renovated	Year of the most recent renovation
sqft_living15	Adjusted square foot of living (Assumed)

sqft_lot15	Adjusted square foot of lot (Assumed)
Coded Variable/Factors	Description
age	Years between 2020 and the highest of yr_built and yr_renovated
cluster	Cluster of the listing derived from KMeans Clustering with k = 3 on price, bedrooms, bathrooms, sqft_living, sqft_lot, floors, sqft_above, sqft_basement, yr_built, yr_renovated, sqft_living15, sqft_lot15, age