

Achievement 6 Project

Brief: Airbnb In NYC

Analysis

Objective

Discover the patterns around Airbnb apartments in NYC.

Data Summary

Data Source:

Open-source data from Kaggle.com

<https://www.kaggle.com/datasets/thedevastator/airbnbs-nyc-overview/>,

originating from <https://huggingface.co/datasets/gradio/NYC-Airbnb-Open-Data>

Data Contents:

The data offers descriptions, rates, reviews and availability for over 20,000 Airbnbs in NYC.

Data Limitations:

It is important this data refers to Airbnb and not apartment rental in NYC in general.

Project motivation

Modern services like deliveries or rentals coming from digitalization are very attractive topics for customers and for data analytics specialists. As customers it is of interest to be aware of new trends and opportunities. As data analyst taking a Airbnb data and conducting analysis on it gives the insights into modern fields and opens the doors into modern data-oriented companies.

Data Profile

Data Cleaning:

Two columns with mixed data types were corrected.

Data Types:

	Data types			
Variable	Time-variant / -invariant	Structured / Unstructured	Qualitative / Quantitative	Qualitative: Nominal / Ordinal Quantitative: Discrete / Continuous
id	time-invariant	structured	quantitative	discrete
name	time-invariant	unstructured	qualitative	nominal
host_id	time-invariant	structured	quantitative	discrete
host_name	time-invariant	unstructured	qualitative	nominal
neighbourhood_group	time-invariant	unstructured	qualitative	nominal
neighbourhood	time-invariant	unstructured	qualitative	nominal
latitude	time-invariant	structured	quantitative	continuous
longitude	time-invariant	structured	quantitative	continuous
room_type	time-invariant	structured	qualitative	ordinal
price	time-variant	structured	quantitative	continuous
minimum_nights	time-variant	structured	quantitative	discrete
number_of_reviews	time-variant	structured	quantitative	discrete
last_review	time-variant	unstructured	qualitative	nominal
reviews_per_month	time-variant	structured	quantitative	continuous
calculated_host_listings_count	time-variant	structured	quantitative	discrete
availability_365	time-variant	structured	quantitative	discrete

Descriptive Statistics:

	id	host_id	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
count	48895	48895	48895	48895	48895	48895	48895	38843	48895	48895
mean	19017143	67620011	40.72895	-73.9522	152.7207	7.0299622	23.274466	1.37322143	7.143982	112.78133
std	10983108	78610967	0.05453	0.046157	240.1542	20.51055	44.550582	1.680442	32.952519	131.62229
min	2539	2438	40.49979	-74.2444	0	1	0	0.01	1	0
25%	9471945	7822033	40.6901	-73.9831	69	1	1	0.19	1	0
50%	19677284	30793816	40.72307	-73.9557	106	3	5	0.72	1	45
75%	29152179	1.07E+08	40.76312	-73.9363	175	5	24	2.02	2	227
max	36487245	2.74E+08	40.91306	-73.713	10000	1250	629	58.5	327	365

Data Ethics:

Data includes no PII. There could be collection (assuming hosts need to consent gathering the data) and measurement (rounding up or down) bias.

Key Questions

- Where in NYC the most apartments for rent are situated and what is the price distribution?
- Which factors other than location influence the price?
- Are there apartment profiles?
- Who are the top Airbnb hosts in NYC?