Metadata

Newly Created Data - Dataset Level Metadata

Data File(s)	Description
airbnb_subset	A subset of Airbnb listing data. Include relevant variables of short-term rent properties with 2 bedrooms in NYC
zillow_subset	A subset of Zillow data. Include relevant variables of median house price of properties in NYC
dt_full	A joined table of airbnb_subset and zillow_subset by zipcode. Include all columns of airbnb_subset and zillow_subset
dt	A subset of dt_full including the median house value at the scraped date. Price is re-calculated
dt_top	A subset of dt including the top 10 zipcodes with highest score. It's an intermediate dataset for drawing investment map
neighbor_summary	A dataset generated by dt_full and grouped by neighborhood. It includes the average house values of each neighborhood in each date.
t_neighbor_summary	A time series dataset transformed by neighbor_summary. It's prepared for the visualization of trend.
zip_summary	A dataset including average number, monthly/annual revenue, cost, return ratio, trend, and score grouped by zipcode
zillow_add	The lastest version of Zillow data. Include historical estimated value for two-bedroom properties from 1996-04 to 2019-12.

Column Level Metadata

airbnb_subset

Field	Description
id	Identifier used by AirBnB for the listing
last_scraped	Identifier used internally by AirBnB to identify when the data was last pulled.
neighbourhood	Neighborhood where the property is located
neighbourhood_cleansed	Verified neighborhood name where the property is located.
neighbourhood_group_cleansed	Name of the area where the property is located.
zipcode	Zip code where the property is located.
latitude	The angular distance of a place north or south of the earth's equator, expressed in degrees and minutes.

longitude	The angular distance of a place east or west of the meridian at Greenwich, England, expressed in degrees and minutes.
room_type	Indicates specific rooms available for rent and/or the entire home.
square_feet	Square footage of the property or space for rent.
price	Price the host is charging to stay per night.
weekly_price	Weekly price the host is charging; this could be discounted for longer term stays. Missing values are replaced by 7 times of price.
monthly_price	Monthly price the host is charging; this could be discounted for longer term stays. Missing values are replaced by 30 times of price.
minimum_nights	Minimum amount of nights the host is willing to rent out the property.
maximum_nights	Maximum amount of nights the host is willing to rent out the property.

zillow_subset

Field	**Description**
RegionID	Zillow assigned number only, assigned consecutively when the regions are defined.
zipcode	= RegionName Zip code of where the property is located.
1996-04 – 2017-06	Indicates the historical median price within that area.
intercept	The intercept generated by the linear regression model of historical data to fit the trend of house value
slope	The slope generated by the linear regression model of historical data to fit the trend of house value
scraped_date	The house value at scraped date generated by ARIMA forecasting

dt

Field	**Description**
Columns in airbnb_subset	All columns in airbnb_subset, from zipcode to maximum_nights
RegionID	Zillow assigned number only, assigned consecutively when the regions are defined.
scraped_date	The house value at scraped date generated by ARIMA forecasting
growth_rate	The slope generated by the linear regression model of historical data to fit the trend of house value
average_daily_price	Average daily price calculated by a metrics of conditions. If minimum_nights <7 & maximum_nights<7){ avg_daily_price = price

	} else if (minimum_nights<7 & maximum_nights>=7 & maximum_nights<30){
	avg_daily_price = daily_prob*price
	+(weekly_prob+monthly_prob)*weekly_price/7
	} else if (minimum_nights<7 & maximum_nights>=30){
	avg_daily_price =
	daily_prob*price+weekly_prob*weekly_price/7+
	monthly_prob*monthly_price/30
	} else if (minimum_nights>=7 & minimum_nights<30
	& maximum_nights>=7 &
	maximum_nights<30){
	avg_daily_price = weekly_price/7
	} else if (minimum_nights>=7 & minimum_nights<30
	& maximum_nights>=30){
	avg_daily_price =
	(daily_prob+weekly_prob)*weekly_price/7
	+monthly_prob*monthly_price/30
	} else if (minimum_nights>=30 &
	maximum_nights>=30){
	avg_daily_price = monthly_price/30

monthly_rev	= average_daily_price * 30 * occupancy rate
annual_rev	= monthly_rev * 12

t_neighbor_summary

Field	Description
date	Date from 1996-04 to 2017-06
Brooklyn	Average house price of each zipcode in Brooklyn
Manhattan	Average house price of each zipcode in Manhattan
Queens	Average house price of each zipcode in Queens
Staten Island	Average house price of each zipcode in Staten Island

zip_summary

Field	Description
zipcode	Date from 1996-04 to 2017-06
count	Number of properties of each zipcode
avg_monthly_rev	Average monthly revenue of properties of each zipcode
avg_annual_rev	Average annual revenue of properties of each zipcode
avg_cost	Average cost of properties of each zipcode
payback_year	= avg_cost/avg_annual_rev Average payback year of properties of each zipcode
rev_cost_ratio	= avg_annual_rev/avg_cost Average revenue to cost ratio of properties of each zipcode
trend	Average growth rate of properties of each zipcode
neighborhood	= neighbourhood_group_cleansed

	Name of the area where the property is located.
score	The weighted average of normalized key factors. = weight_count * (count / mean(count)) + weight_cost * (1 / (avg_cost / mean(avg_cost)))+ weight_rev * (avg_annual_rev / mean(avg_annual_rev)) + weight_payback * (1 / (payback_year / mean(payback_year))) + weight_profit * (rev_cost_ratio / mean(rev_cost_ratio)) + weight_trend * (trend / mean(trend))