T-I **AIR BNB**

What we can draw from the data to find

Trends in Airbnb Rental Calendar

T-I

BUSINESS PROBLEM & IMPACT

AirBnB was created to answer the question of affording rent in an expensive city. Hosts are able to subletting their homes when they are not using it

By identifying trends in the demand and prices, the company will be able to develop:

- Pricing strategies and marketing initiatives
- Optimal marketing plans and outreach initiatives
- Competitor analysis

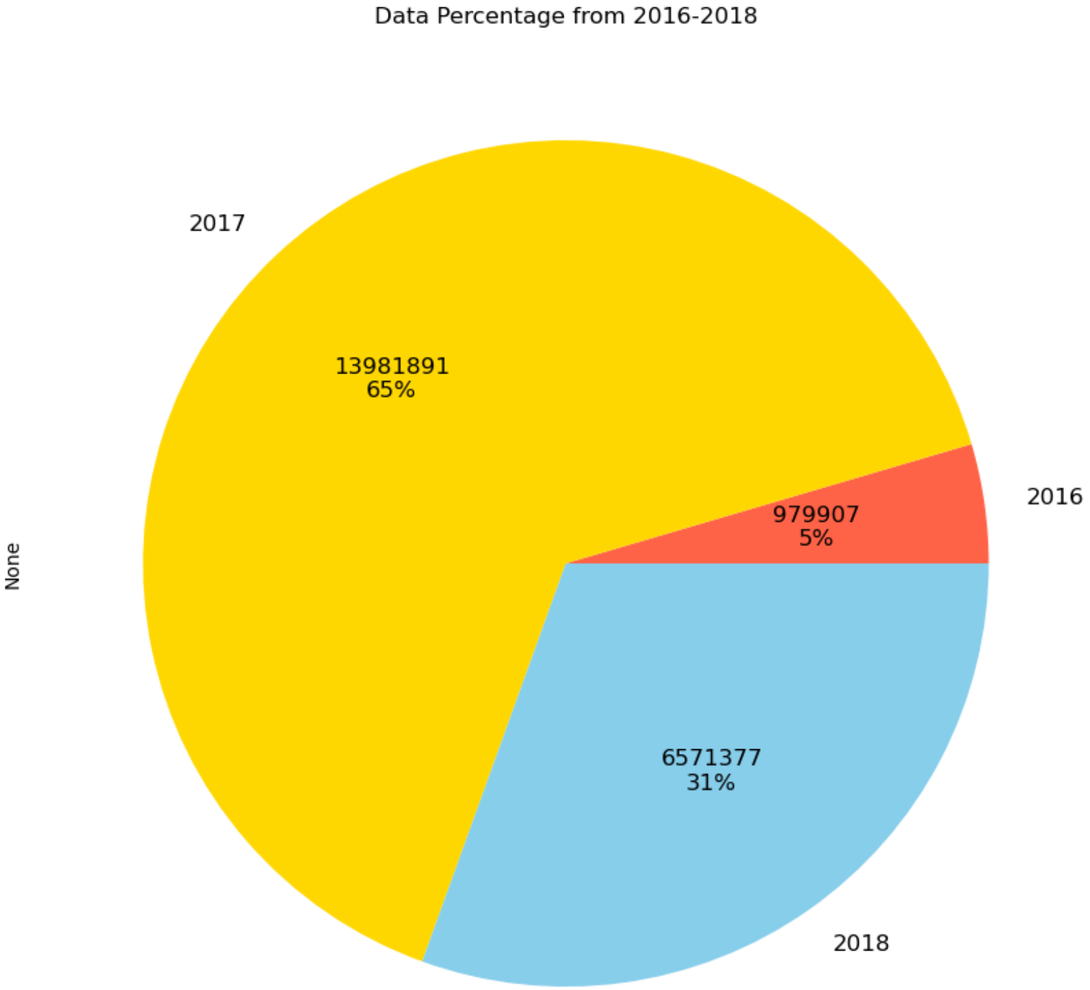
Table of Contents

PART 1

Overall pictures

T-I

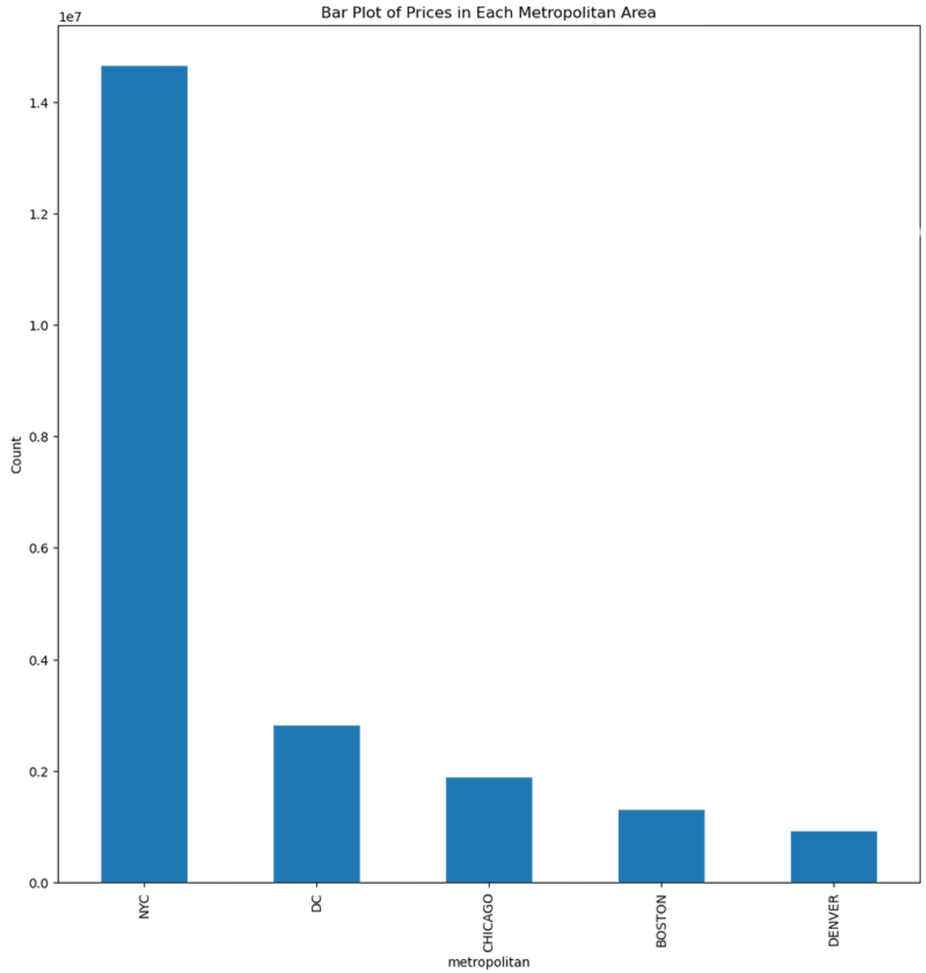
THE OVERALL PICTURE OF THE DATA



Majority of the data is from 2017

T-I

GEOGRAPHICAL COMPONENTS

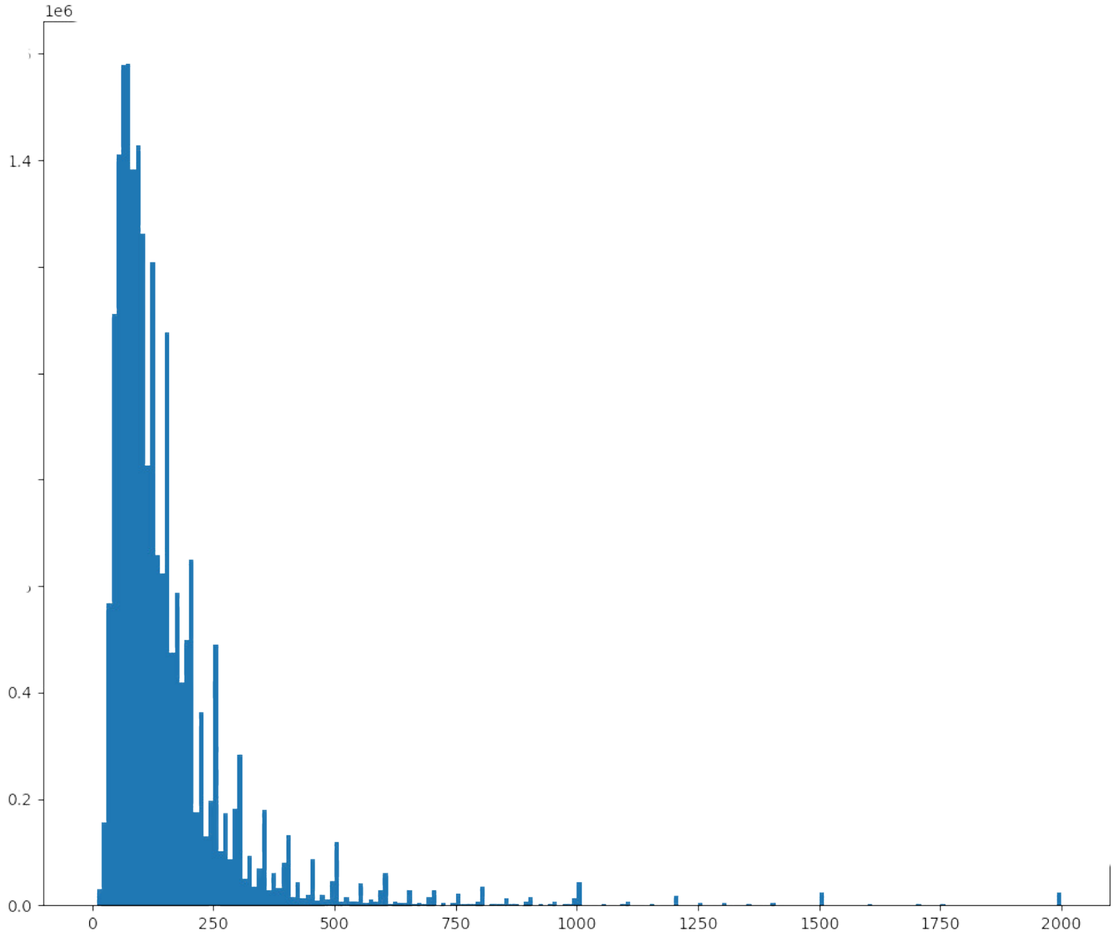


Ranking form Highest to Lowest

- NYC
- DC
- Chicago
- Boston
- Denver

T-I

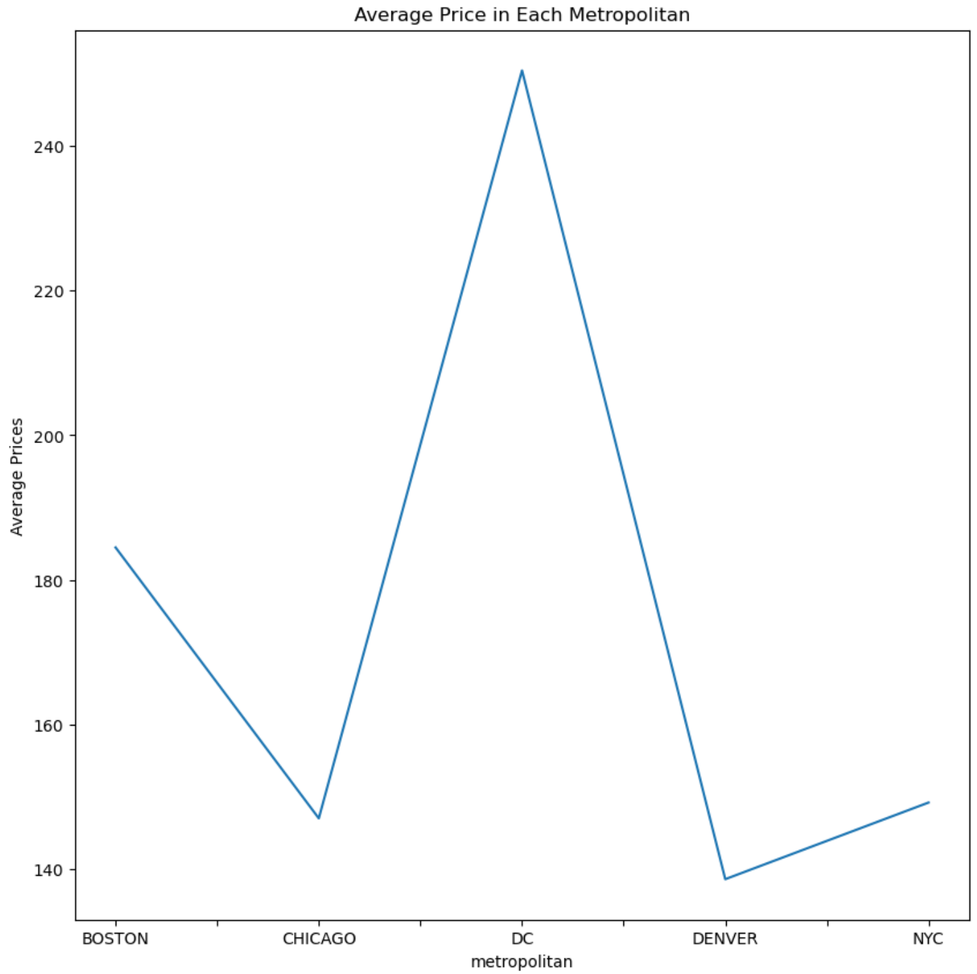
NIGHTLY RATES DISTRIBUTION



Nightly rates revolve around 0 to 200 dollars with outliers that are greater than 2000 dollars.

T-I

THE MOST EXPENSIVE METROPOLITAN AREA ON AVERAGE.



Average price ranking:

1. DC
2. Boston
3. Chicago
4. NYC
5. Denver

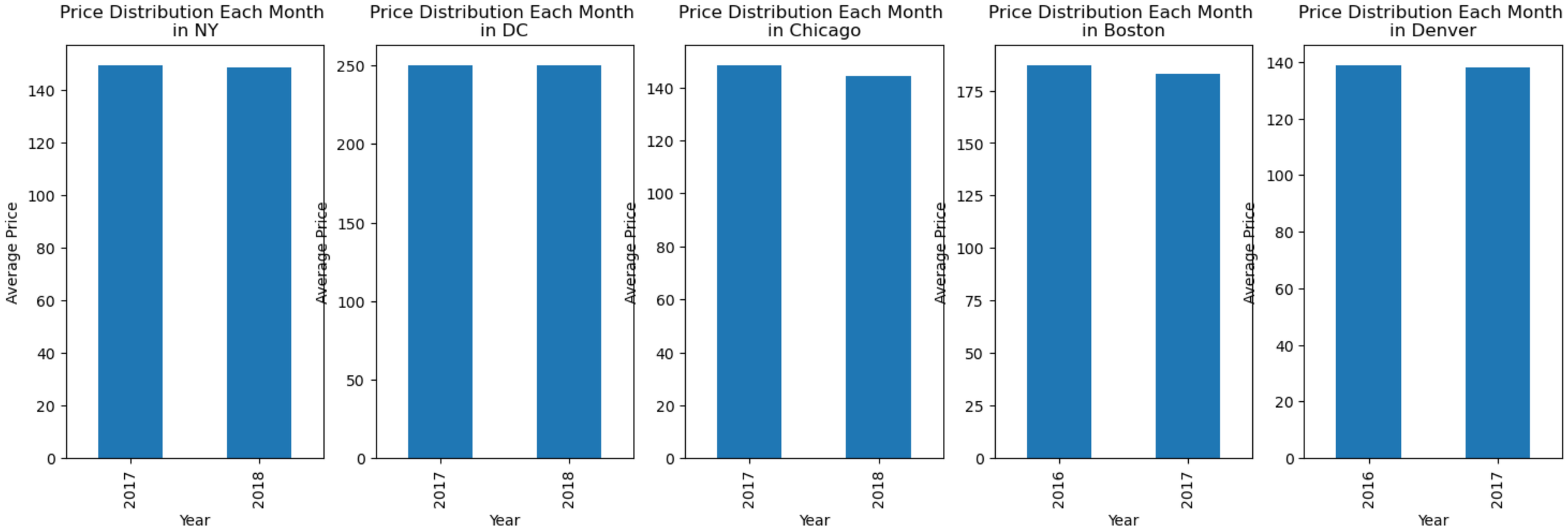
PART 2

Analysis on each Metropolitan

T-I

PRICE DISTRIBUTIONS BY YEAR

VERY LITTLE CHANGE IN THE PRICE.



T-I

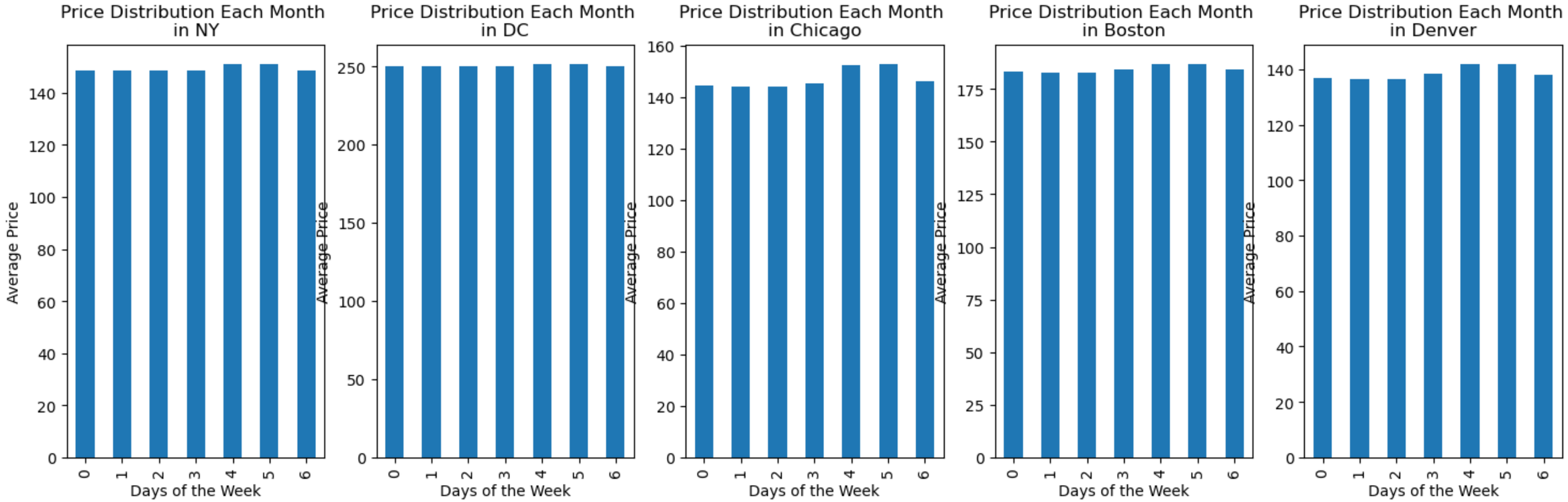
PRICE DISTRIBUTIONS BY MONTH

SHOW SOME SORT OF CYCLICAL TREND

T-I

PRICE DISTRIBUTIONS BY DAY OF THE WEEK

PRETTY STABLE, NO CLEAR TREND DISCOVERED



Explore

Neighborhoods in each Metropolitan

TO FIND

Trends over time

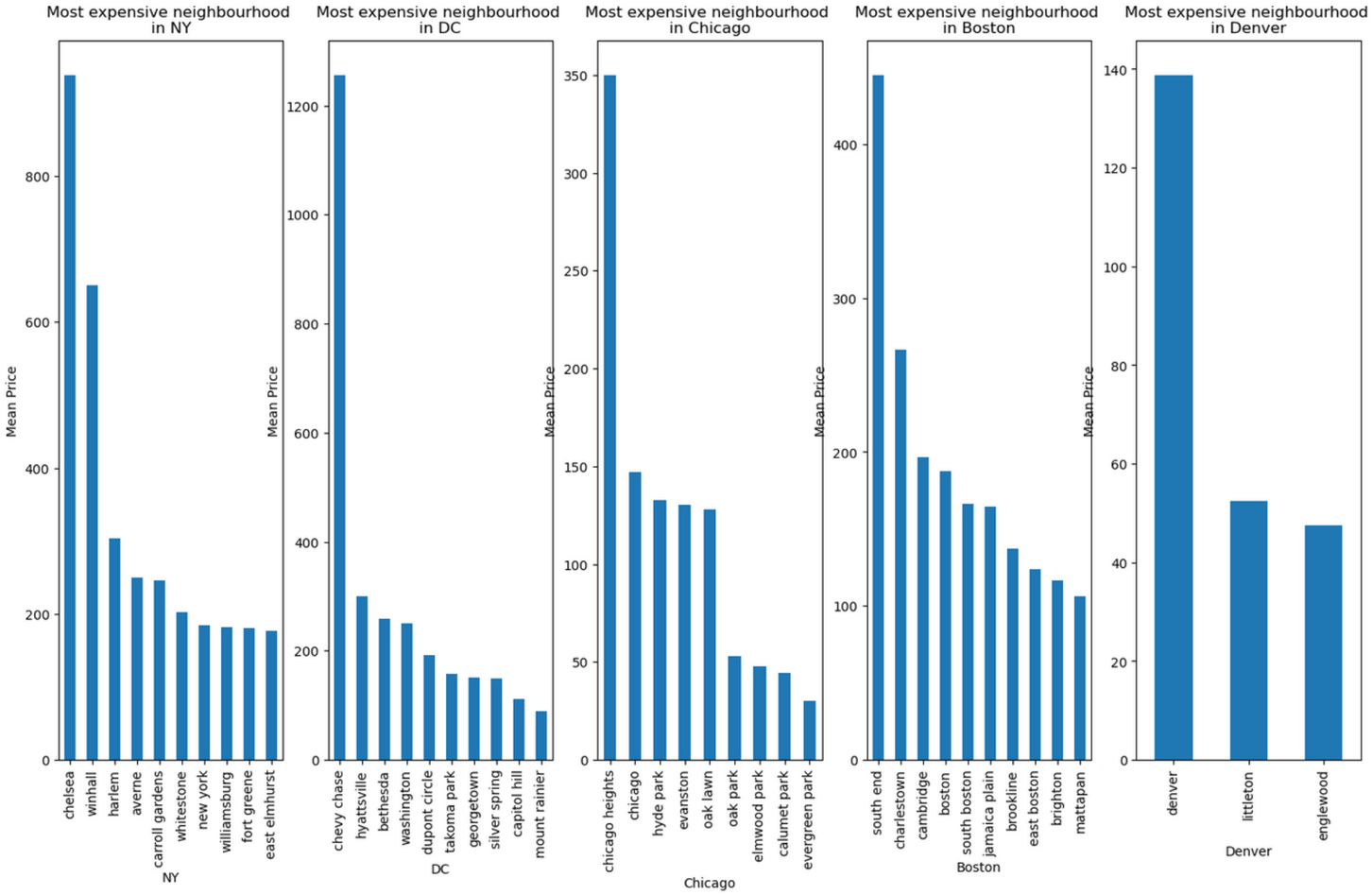
can their prices be justified?

THE ANALYSIS OF

The most Expensive

T-I

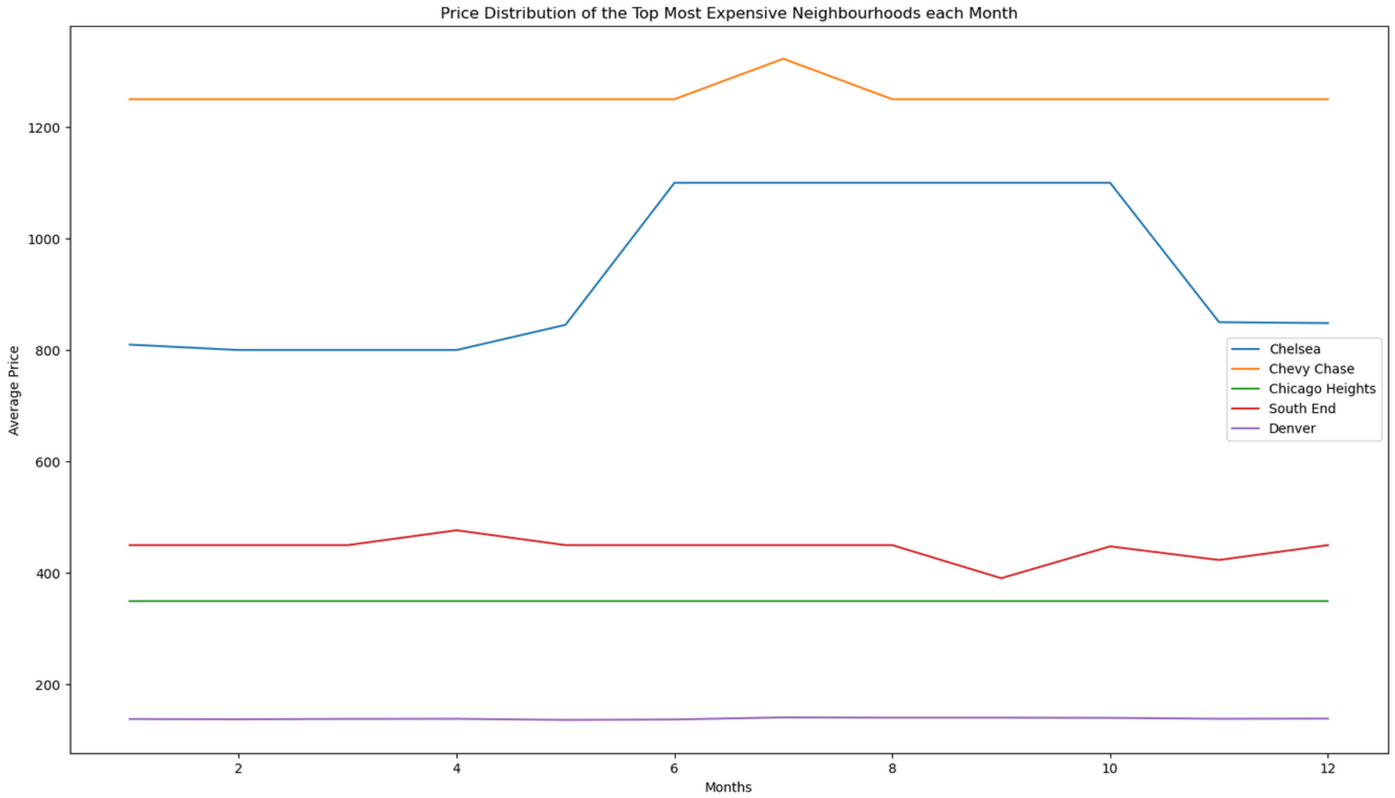
THE MOST EXPENSIVE NEIGHBOURHOODS



- **New York:** Chelsea
- **DC:** Chevy Chase
- **Chicago:** Chicago Heights
- **Boston:** South End
- **Denver:** Denver

T-I

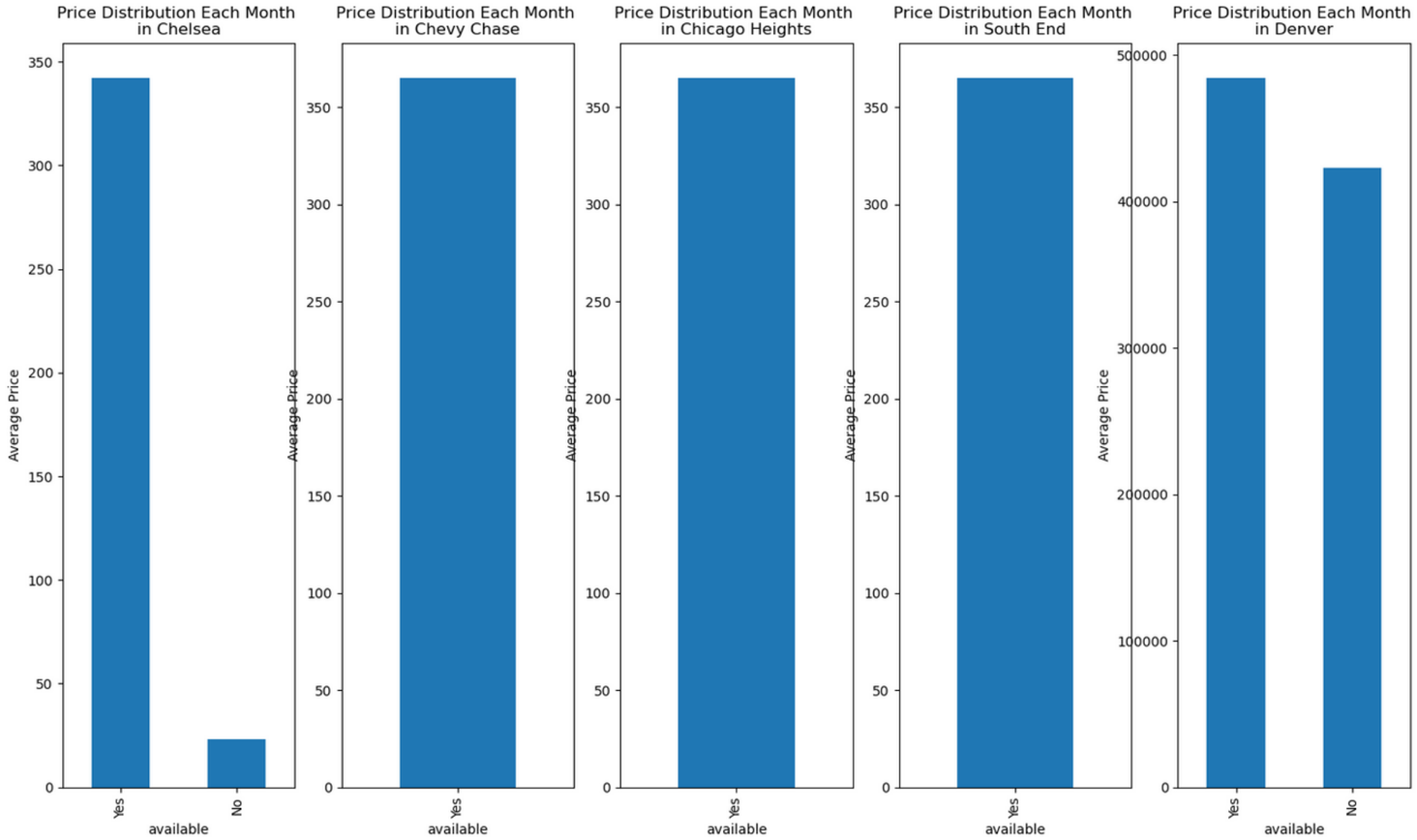
PRICE DISTRIBUTION OF THE TOP MOST EXPENSIVE NEIGHBOURHOODS EACH MONTH



Only in **Chelsea & NY**: there's a huge change in prices during the summer, otherwise, we don’t see much difference in price.

T-I

AVAILABILITY

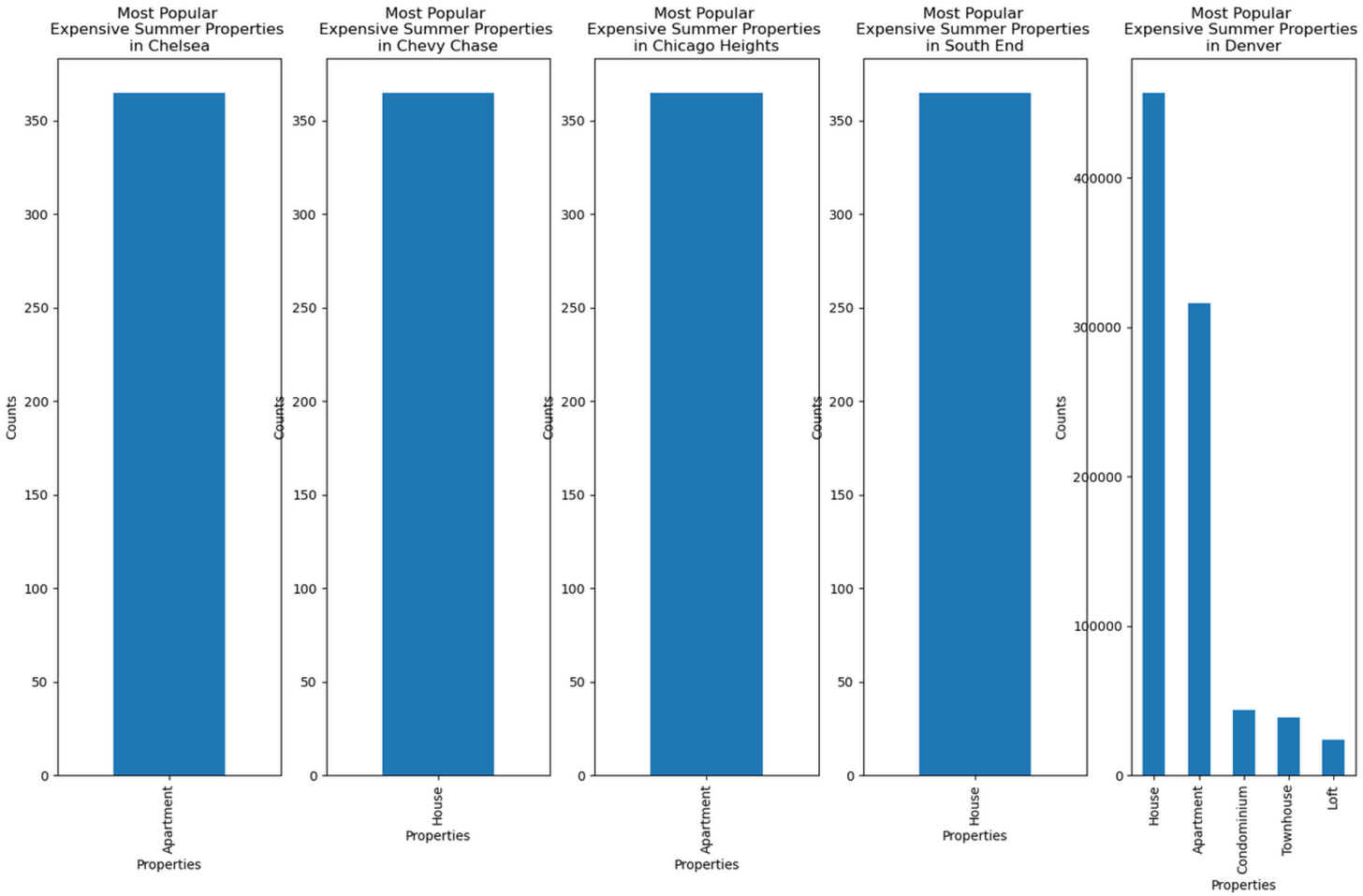


Almost every neighborhood but **Denver** is always available

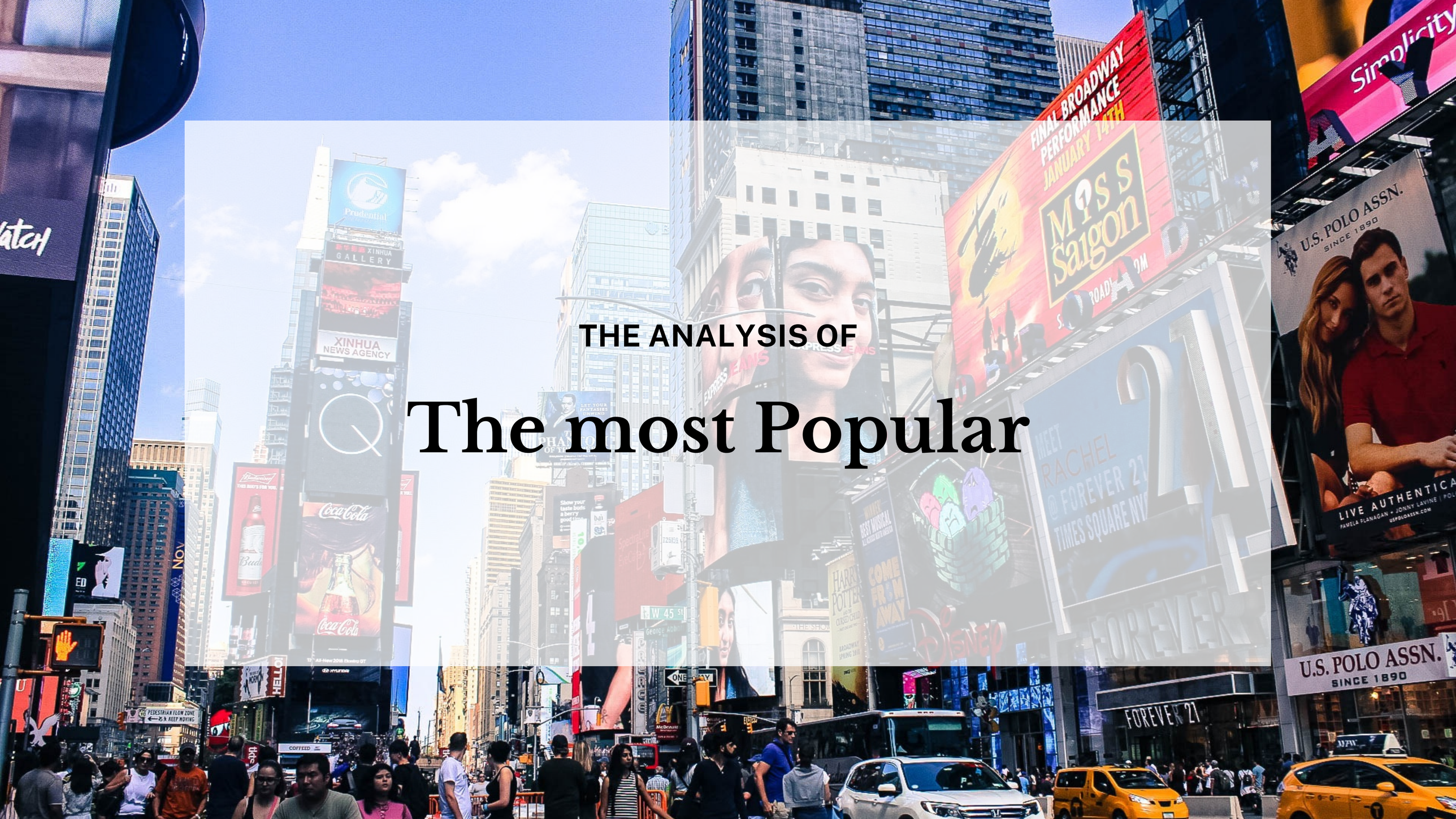
This tells us that these properties don't have much demand.

T-I

THE TYPE OF PROPERTY

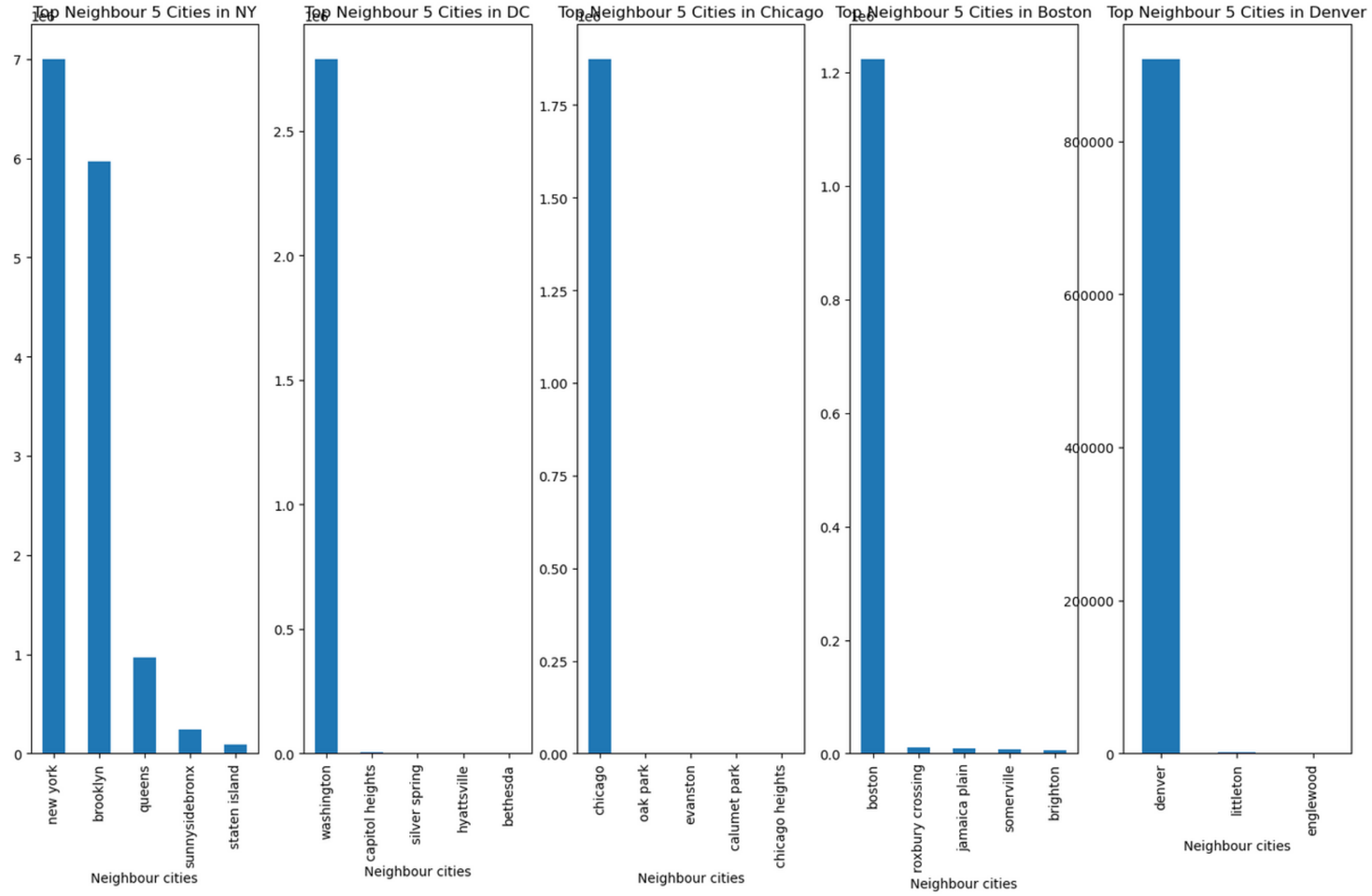


Among all the property types **Houses & apartments** are the most popular


THE ANALYSIS OF
The most Popular

T-I

TOP 5 POPULAR NEIGHBOURHOODS

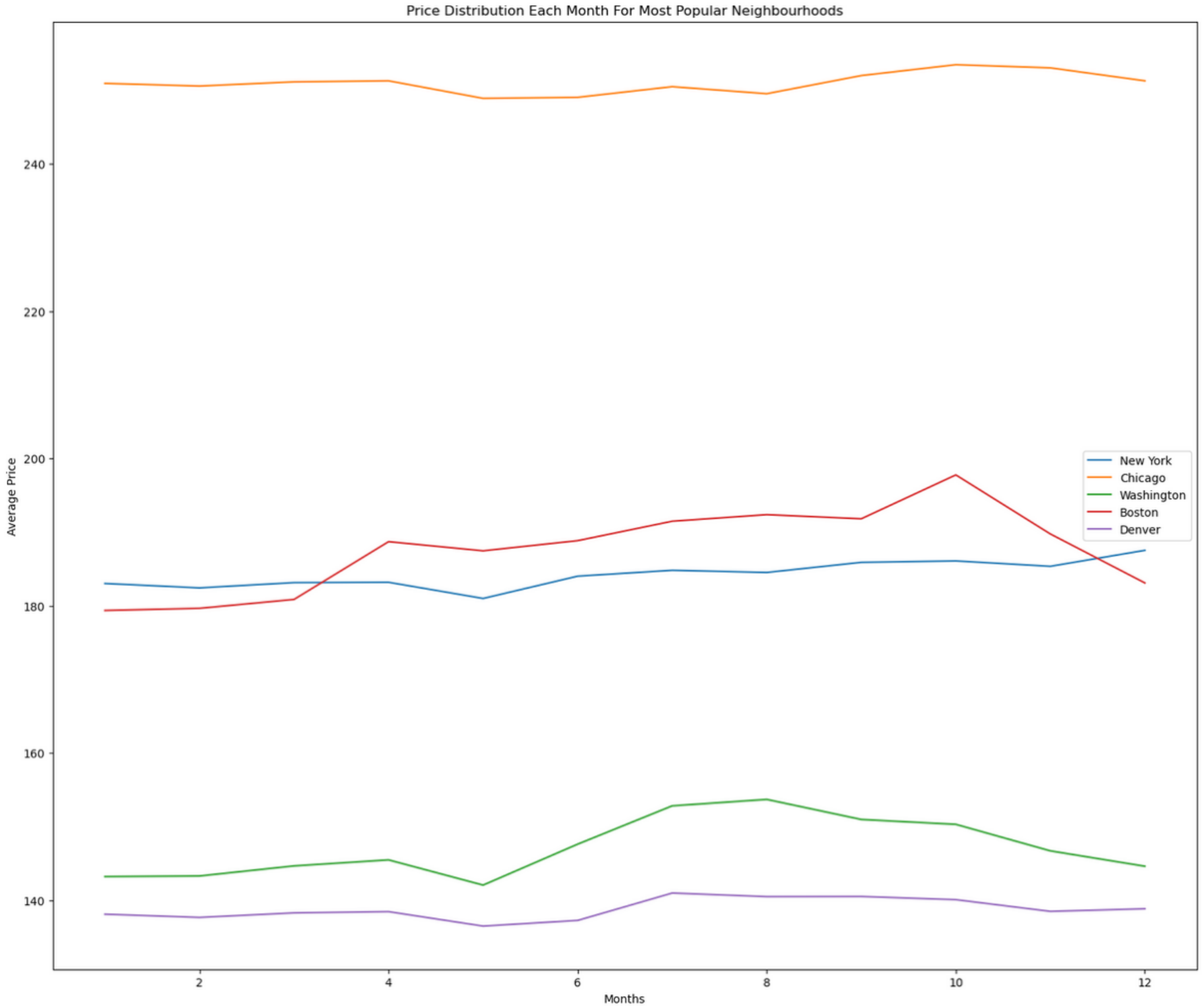


The city names of the metropolitan areas are the most popular spots:

- New York
- DC
- Chicago
- Boston
- Denver

T-I

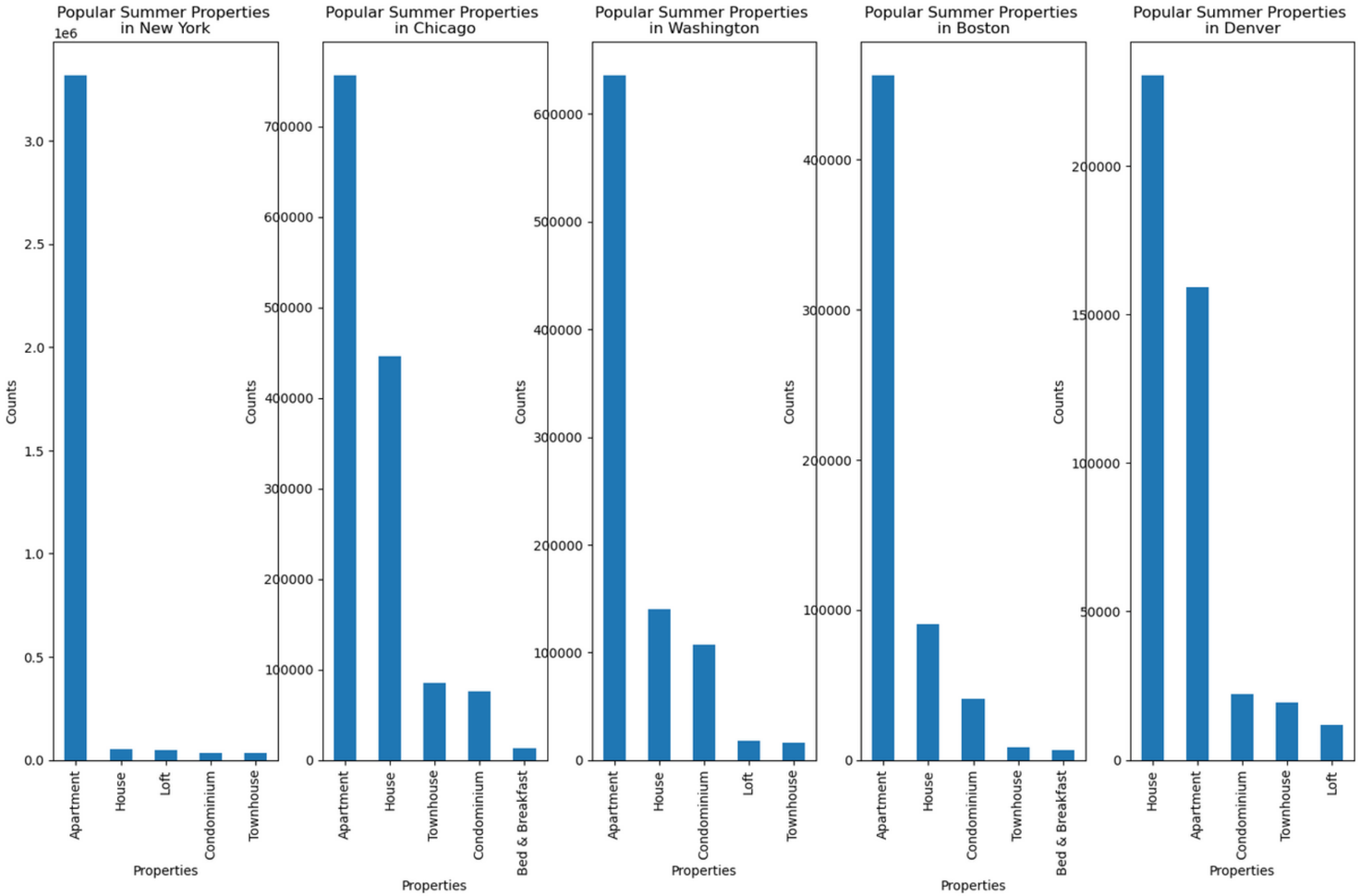
PRICE DISTRIBUTION EACH MONTH FOR MOST POPULAR NEIGHBOURHOODS



Almost all of the cities, there is a spike during the summer period (May till August/September) **vs.** the rest of the months of the years

T-I

THE TYPE OF PROPERTY

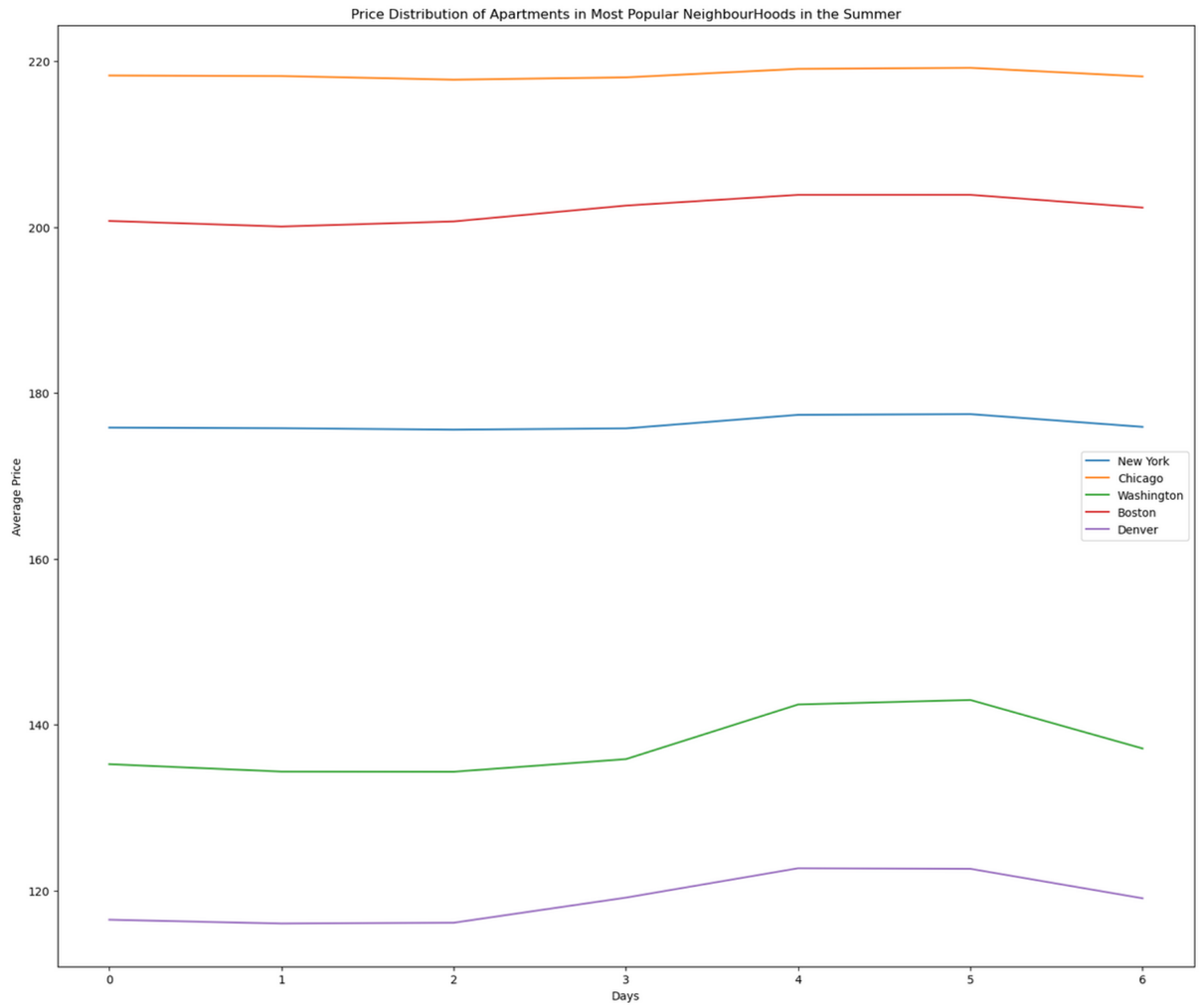


Apartments are the most popular during the summer

But for Denver which would be **houses**

T-I

PRICE DISTRIBUTION OF APARTMENTS IN MOST POPULAR NEIGHBOURHOODS IN THE SUMMER



Friday & Saturday are when the prices increase for the popular city, and then fall back down on **Sunday**.

PART 3

Regression Model

Multiple Regression

Build a model, using the previous knowledge of trends we found

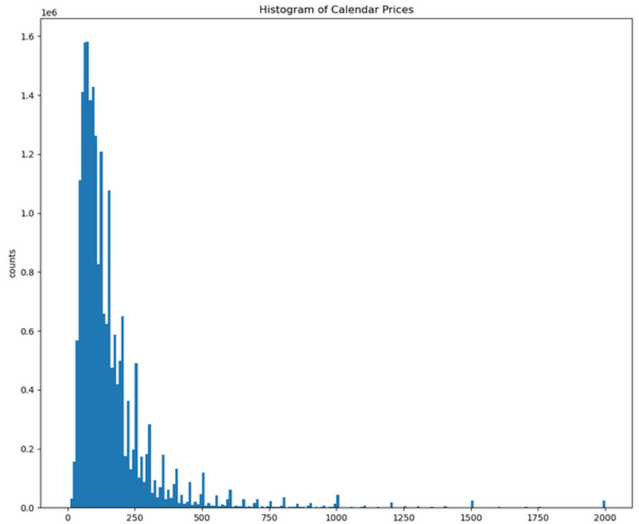
T-I

MULTIPLE REGRESSION

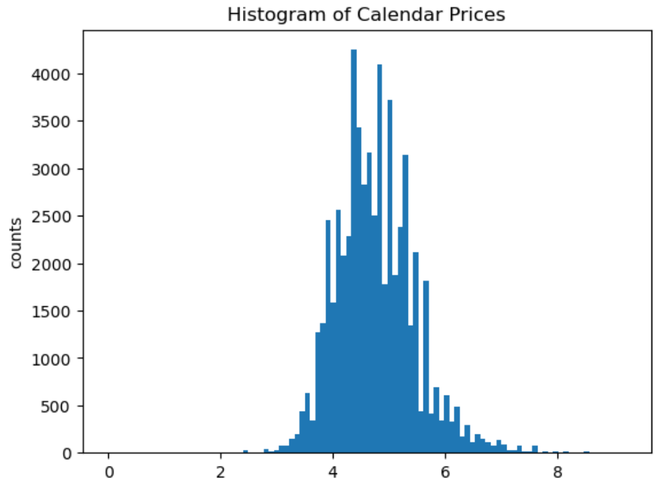
We **firmly believe** that **the most simple model** is the only model required for the listings dataset

The assumption:

The error term $\varepsilon \sim N(0, \sigma^2)$ and are independent of any of the covariates we define



Prices are really skewed to the right.
Hence we can apply a log transformation, to make it more normally distributed



T-I

MULTIPLE REGRESSION

The variables we use for the model that we believe are **the best predictors** to explain the prices are as follows:

- accommodates
- bed type
- bedrooms
- city
- instant bookable
- metropolitan
- property type
- review scores location
- review scores value
- room type

Our prediction model is:

$$Price = \beta_0 + \beta_1 \times accommodates + \beta_2 \times bedtype + \cdots + \beta_9 \times roomtype$$

MULTIPLE REGRESSION

OLS Regression Results

Dep. Variable:	price_log	R-squared:	0.592
Model:	OLS	Adj. R-squared:	0.590
Method:	Least Squares	F-statistic:	341.2
Date:	Mon, 10 Apr 2023	Prob (F-statistic):	0.00
Time:	07:23:13	Log-Likelihood:	-29563.
No. Observations:	47195	AIC:	5.953e+04
Df Residuals:	46994	BIC:	6.129e+04
Df Model:	200		
Covariance Type:	nonrobust		

$R^2 \approx 0.6$ essentially tells us 60% of the variation is explained by the covariates we have chosen.

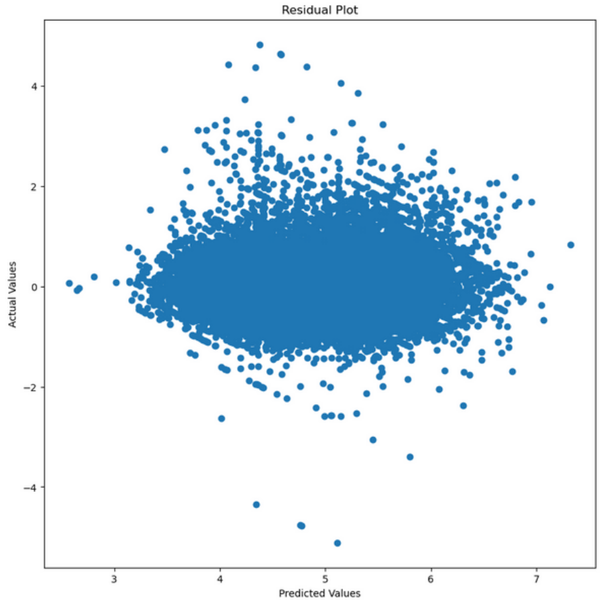
The model summary, with critical value α = 0.05, gives the following covariates to be considered statistically significant:

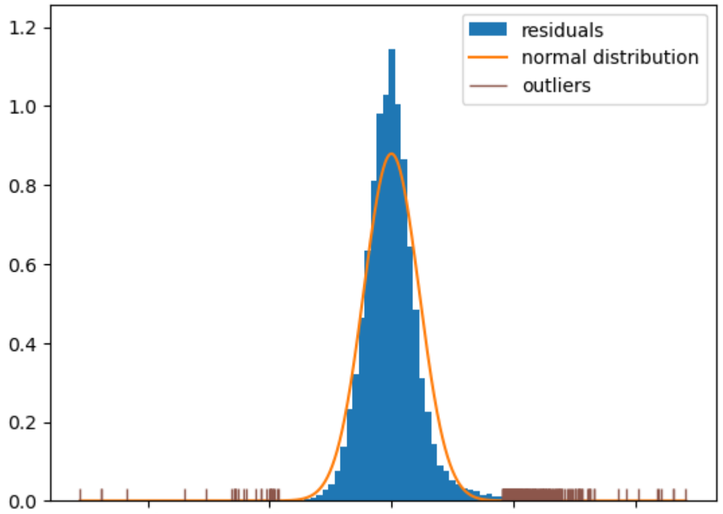
- availability 30
- room type
- accommodates
- bedrooms
- property type
- city
- review scores location

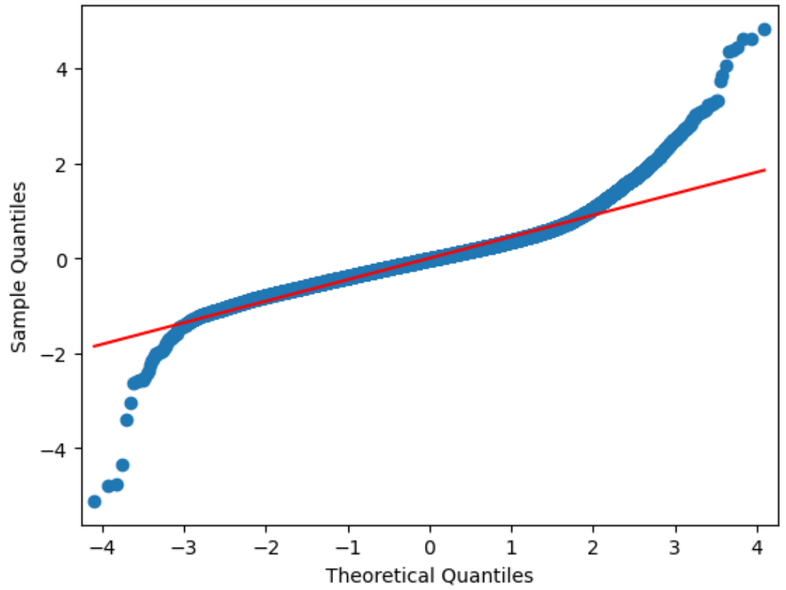
T-I

MULTIPLE REGRESSION

Assumptions testing:







Good signs that the residuals do not get out of hand
The QQ plot indicates that our errors are normal and linear.

Conclusion

BUSINESS QUESTION	OUR FIRST STEP	OUR FOCUS	DEVELOPED A PREDICTION MODEL
Identify trends in the Airbnb rental calendar over time	Process the data	Discover useful trends in neighbourhoods level factors	Using a multiple regression model with an outcome of $R^2 \approx 0.6$
How these might be explained by listing-specific and/or neighborhood-level factors	Keeping valuable information, and not being misled by incorrect data inputs.	In the most expensive neighborhood, price is driven by the location and property type, where 50% is apartment and 50% is houses	Assumptions were not violated at all
To utilize trends to drive up future business and also identify ways to make customers more happy with the use of Airbnb.	Decided to merge the listing and calendar datasets to proceed with our analysi	in the most popular neighborhoods, prices are higher in the summer and especially on Friday and Saturday each week	The model is simple enough for us to use as a prediction of prices for any listing.


Thank you!