



Análisis de algunos* factores
de competitividad para
prestadores de servicio AirBnB
en CDMX

-Aarón García

Objetivo

Dados los alcances de la unidad de aprendizaje, se busca aplicar la mayor cantidad de conocimientos adquiridos a lo largo del curso, centrado en datos abiertos de AirBnB y relacionarlo con datos demográficos, territoriales y de transporte en CDMX, donde al final del ejercicio se busca conocer que dimensiones tienen mayor peso o importancia en el éxito o fracaso de una oferta de AirBnB.

Preguntas de minería a responder

- ¿El número de estaciones de metro y metrobús son un factor importante?
- ¿Qué servicio de transporte es más atractivo?
- ¿Importa el número de hoteles? ¿La categoría (en estrellas) son un factor importante?
- ¿Delegaciones más atractivas?
- ¿Precios pico y promedio por alcaldía, temporada y año?
- ¿El número de noches mínimas importa?
- ¿Distribución de tipo de oferta por alcaldía y temporada del año?
- ¿El número de reviews y el último review son factores importantes?
- ¿Dónde hay más y menos competencia para un oferente?

Entendimiento de los datos



Aarón García <a... mar, 18 may. 00:47 (hace 13 días) ☆ ↩ ⋮
para data ▼

Dear <http://insideairbnb.com>, I'm a college student in Mexico City (CDMX), I'm learning about data mining and one of my projects for this season needs data from some topic, I decided to use data from your site and I ask me if you could share me Mexico CDMX AirBnB Data from 2019 and 2020.

My purpose is only for school, I really hope you consider my request, greetings from Mexico.

Saludos y muchas gracias, excelente día.

"Vive como si fueras a morir mañana, aprende como si fueras a vivir siempre"

Aarón Antonio García González



Murray Cox <murray@insideairbn... 30 may. 2021 11:39 (hace 1 día) ☆ ↩ ⋮
para mí, data ▼

Hi

Inside Airbnb is a mission driven activist project with the objective to:

Provide free data that quantifies the impact of short-term rentals on housing and residential communities; and also provides a platform to support advocacy for policies to protect our cities from the impacts of short-term rentals.

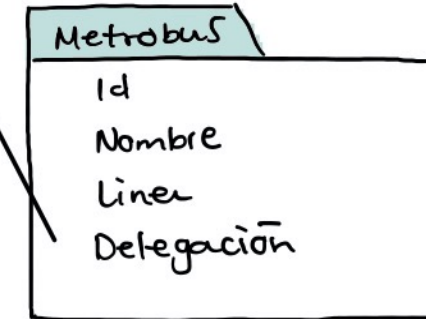
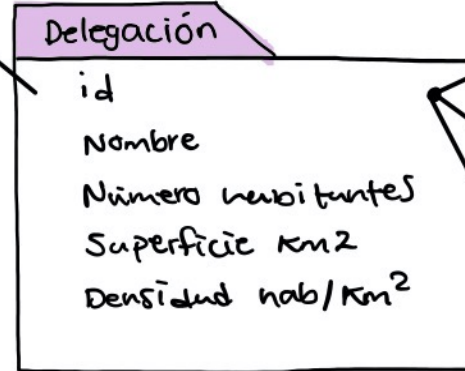
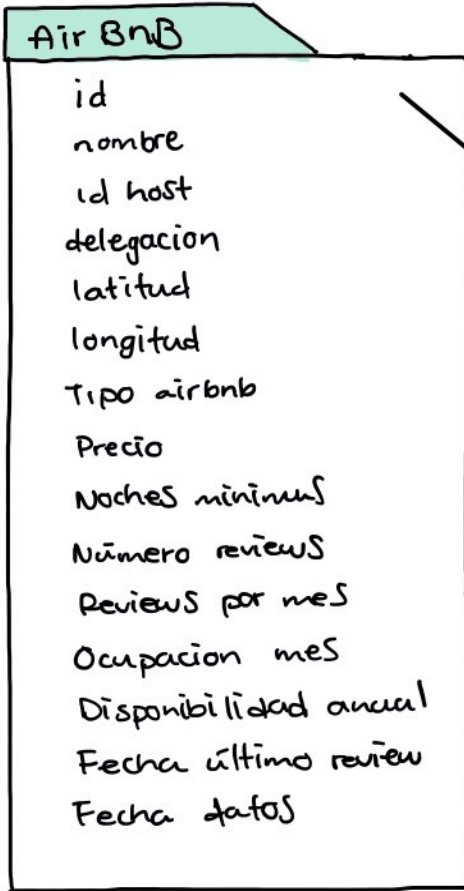
After offering free data across multiple cities and dates for more than 5 years, it's no longer possible to sustain the project by doing so. The new policy is to offer a reasonable amount of free data (last 12 months) that meets the immediate needs of users the project was designed for; and all other data requests will be assessed based on: the amount of data being requested (across time and space); the intended data use; and the identity of the requestor. Requests for a large amount of data and/or of a use outside the mission of the project may be refused or asked to donate to help sustain the project.

Your request (one city) is modest, and your area of study (learning about data mining) does not particularly align with the mission of the project.

Below are links to the data you requested.

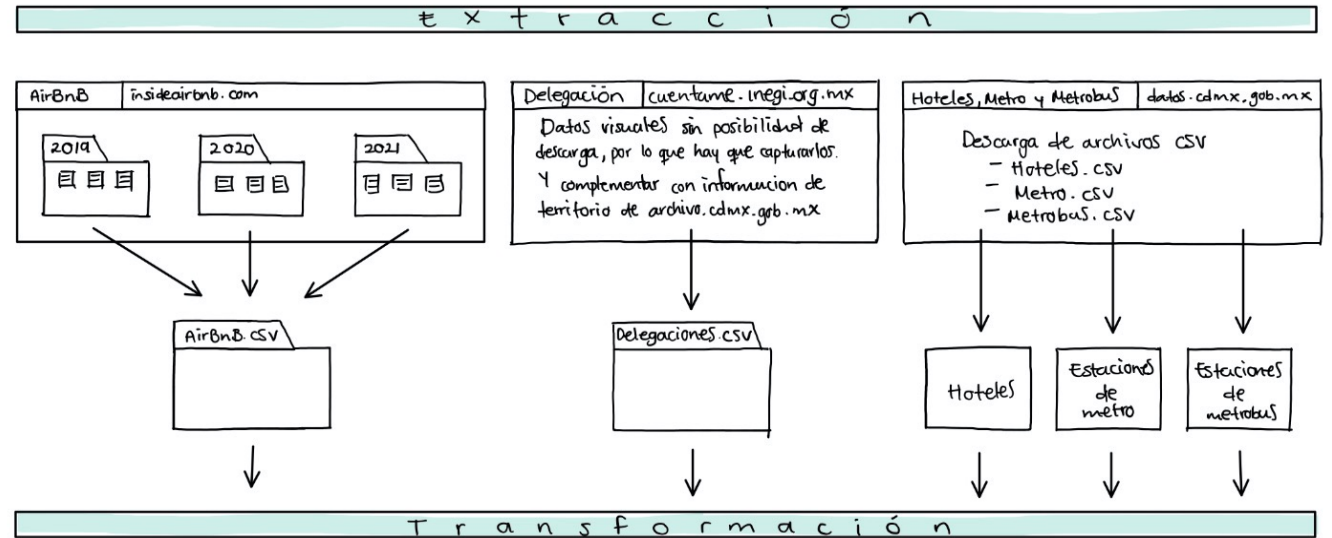
I would also highly encourage you to make some type of [donation](#) to the project to help sustain it for others.

These links expire in 7 days, so please download the data before then.





Microsoft®
SQL Server®



→ Rellenar con fecha comodín a los registros que jamás han sido evaluados.

→ Separar la fecha de último review en día, mes y año, donde únicamente mes y año serán columnas, posterior a ello eliminamos la fecha de review entera del data set.

→ Obtener a partir de cada archivo.csv la fecha de dichos datos, donde el mes y año se agregan como columnas al data set.

→ Eliminar el campo "neighbourhood_group" del dataset

→ Cambiar el nombre de delegación o alcaldía por el id de alcaldía con base al archivo delegaciones.csv.

→ Eliminar registros que incluyan uno o más campos vacíos o no válidos

→ Combinar datos de ambas fuentes de tal manera que podamos identificar a cada delegación y además conocer su nombre, número de habitantes, extensión superficial en kilómetros cuadrados y densidad de población.

→ Para hoteles

→ Eliminar columnas: calle_y_num, colonia y CP

→ Cambiar el nombre de delegación o alcaldía por el id de alcaldía con base al archivo Delegaciones.csv.

→ Para estaciones de metro y metrobús

→ Eliminar columnas: geometry, stop_id, stop_code, stop_desc, stop_lust, stoplon, trip_heads, agency, geopoint.

→ Agregar columna "Delegación", la cual se llena con el id de delegación donde se encuentra la estación, esto se busca estación a estación en la página del servicio.

Carga

Todos los archivos resultantes en el paso de transformación, resultan archivos .csv que son cargados en SQL Server para la minería siguiente.

UNFINISHED (WORKSPACE)
> Desarrollo-de-Sistemas-Distribuidos
 > Data_Mining
 > Practices
 > Project
 > DataSet
 > Images
 > Scripts
 > SQL
 Cleaning.py
 InfoDataSet.py
 script.py
 test.py
 > Tests
 .gitignore
 README.md

aarongarcia@aarongarcia-19-22521a:~/Desktop/Data_Mining/Project/Scripts\$ python3 Cleaning.py

Directorio 2019 :

Procesando 15_03_2019.csv ...
Procesando 24_09_2019.csv ...
Procesando 22_08_2019.csv ...
Procesando 20_10_2019.csv ...
Procesando 25_11_2019.csv ...
Procesando 22_05_2019.csv ...
Procesando 17_04_2019.csv ...
Procesando 16_07_2019.csv ...
Procesando 24_06_2019.csv ...
Procesando 26_12_2019.csv ...

Directorio 2020 :

Procesando 19_03_2020.csv ...
Procesando 26_10_2020.csv ...
Procesando 24_05_2020.csv ...
Procesando 20_06_2020.csv ...
Procesando 27_11_2020.csv ...
Procesando 27_02_2020.csv ...
Procesando 23_12_2020.csv ...
Procesando 23_04_2020.csv ...
Procesando 23_01_2020.csv ...

Directorio 2021 :

Procesando 23_02_2021.csv ...
Procesando 29_01_2021.csv ...
Procesando 22_03_2021.csv ...

Resumen de archivo objetivo

<class 'pandas.core.frame.DataFrame'>

Int64Index: 443241 entries, 0 to 20020

Data columns (total 17 columns):

id	443241	non-null	int64
name	443049	non-null	object
host_id	443241	non-null	int64
neighbourhood	443241	non-null	int64
latitude	443241	non-null	float64
longitude	443241	non-null	float64
room_type	443241	non-null	object
price	443241	non-null	int64
minimum_nights	443241	non-null	int64
number_of_reviews	443241	non-null	int64
reviews_per_month	443241	non-null	float64
calculated_host_listings_count	443241	non-null	int64
availability_365	443241	non-null	int64
month_last_review	443241	non-null	int64
year_last_review	443241	non-null	int64
month_data	443241	non-null	int64
year_data	443241	non-null	int64

dtypes: float64(3), int64(12), object(2)

memory usage: 60.9+ MB

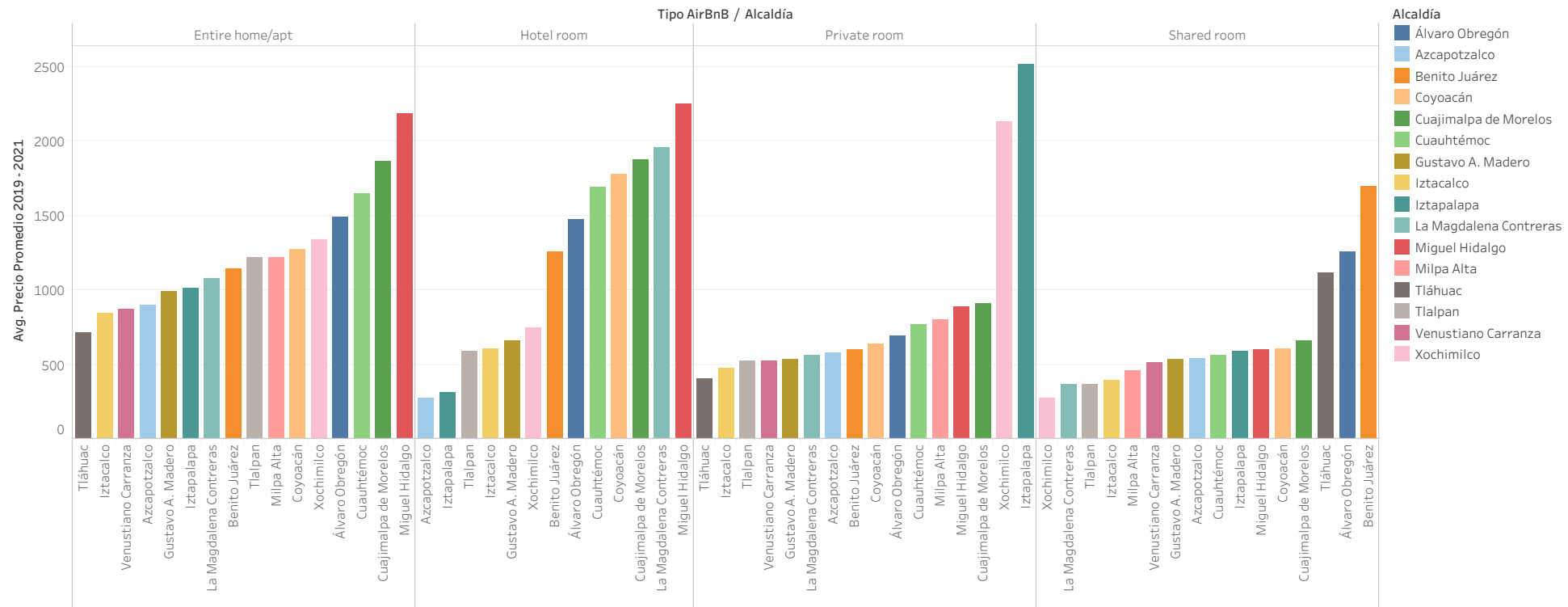
None

Tiempo de procesamiento: 38 segundos

aarongarcia@aarongarcia-19-22521a:~/Desktop/Data_Mining/Project/Scripts\$

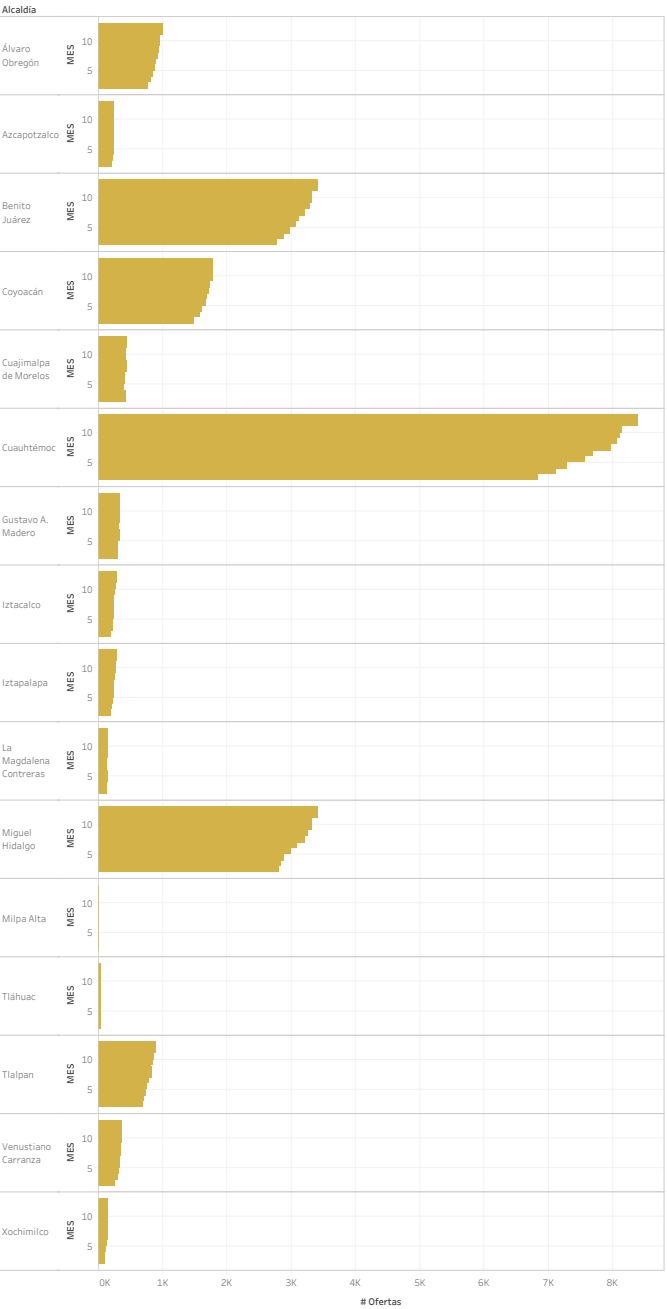
Precio promedio por alcaldía

{Alcaldía}{Tipo de AirBnB}{Precio promedio}

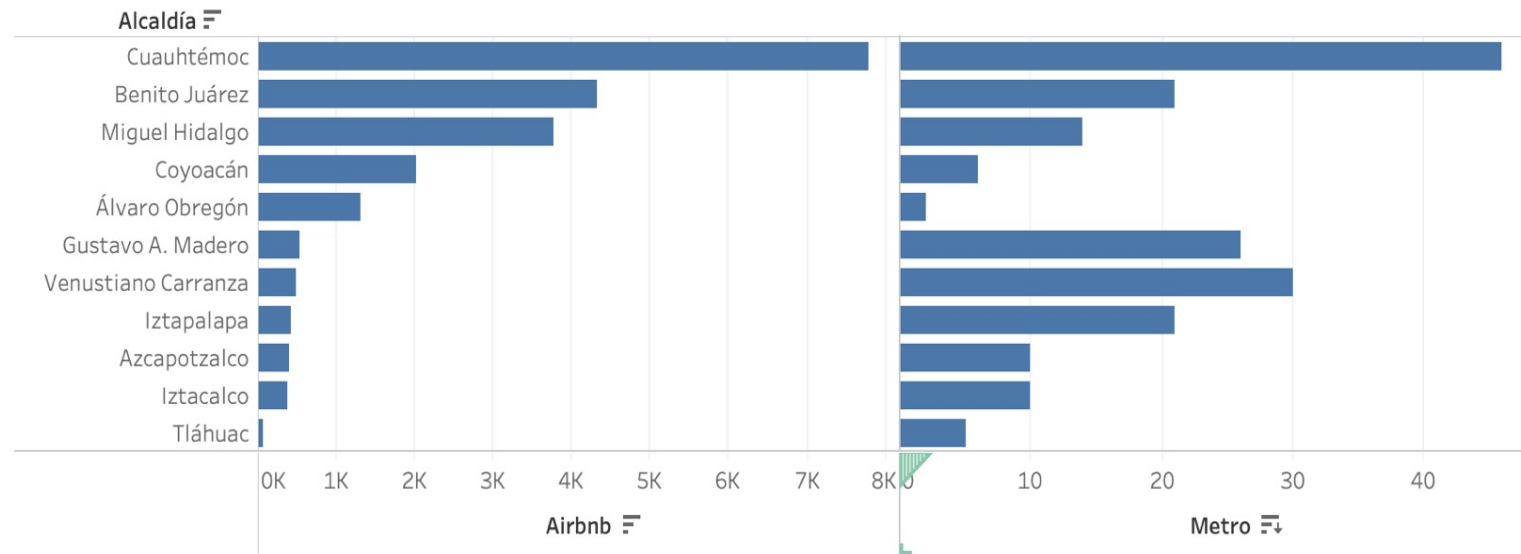


Desde el punto de vista de oferente, es una buena opción ofrecer AirBnB en las delegaciones tales como; Miguel Hidalgo, Cuajimalpa, Cuauhtémoc, Álvaro Obregón, Coyoacán y Xochimilco debido a predominar en precios altos.

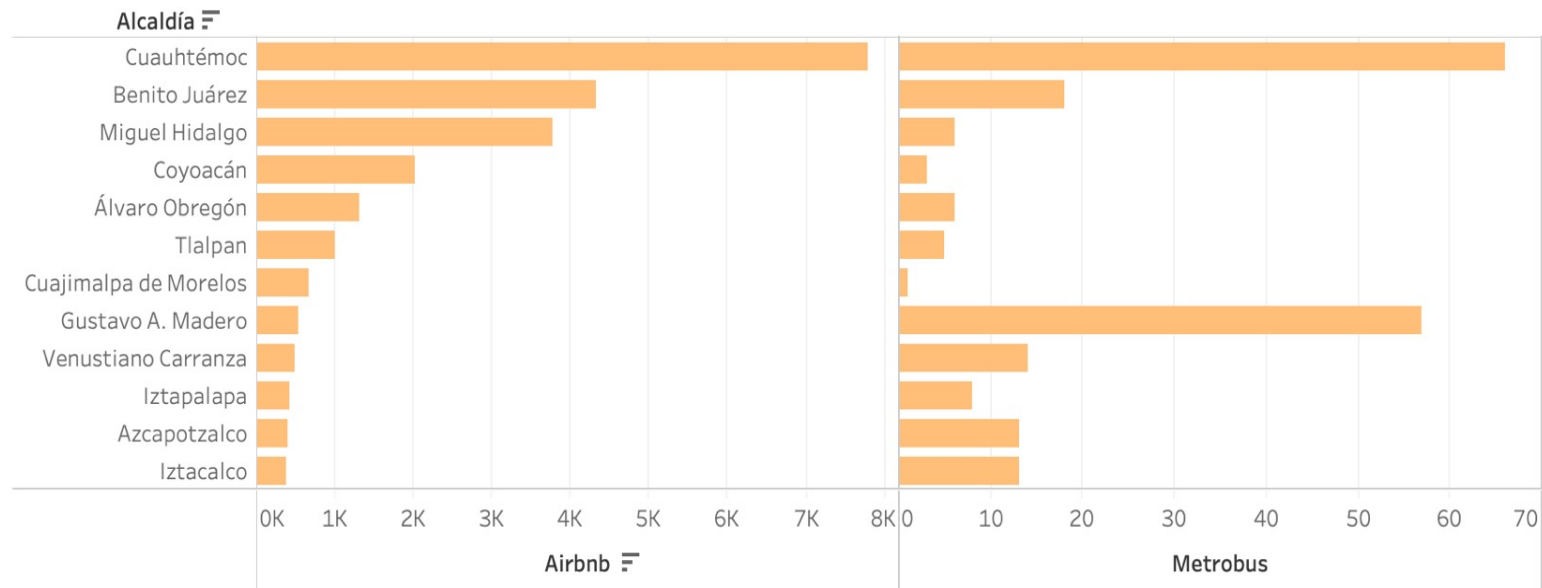
{Aldía}{Numero de AirBnB diferentes}{Mes} 2019



{Alcaldía}{Numero de AirBnB diferentes}{Numero de estaciones de metro}



{Alcaldía}{Numero de AirBnB diferentes}{Numero de estaciones de metrobus}



Un detalle muy importante, la alcaldía Cuauhtémoc siendo la mas atractiva y con mayor oferta, es de las que menor numero de días mínimos de estadía pide, mientras que las otras dos alcaldías mas exitosas en AirBnB (Benito Juárez y Miguel Hidalgo), son de las que mayor numero de días mínimos de estadía piden, además que los precios de estas 3 alcaldías son de los más altos de la ciudad, un contraste muy interesante.