

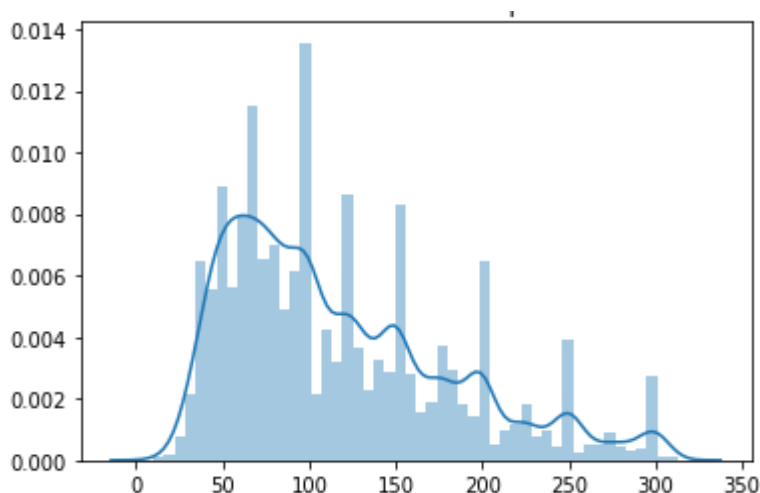
## Reporte Capstone Project

### Descripción del Problema.

Se tiene la información de la Empresa AirBnB del Estado de Nueva York en Estados Unidos para el año 2019 , empresa que provee una aplicación web para que los usuarios puedan poner a disposición sus locales para alquilar a otros usuarios de la misma aplicación web. Se desea analizar cuales atributos ayudan a determinar el precio de alquiler de las localidades disponibles mediante la aplicación.

### Características de la Información Analizada. Principales Hallazgos

Al analizar la información presentada se tienen las siguientes características.



**Gráfico #1: Distribución probabilística de precio en dólares.**

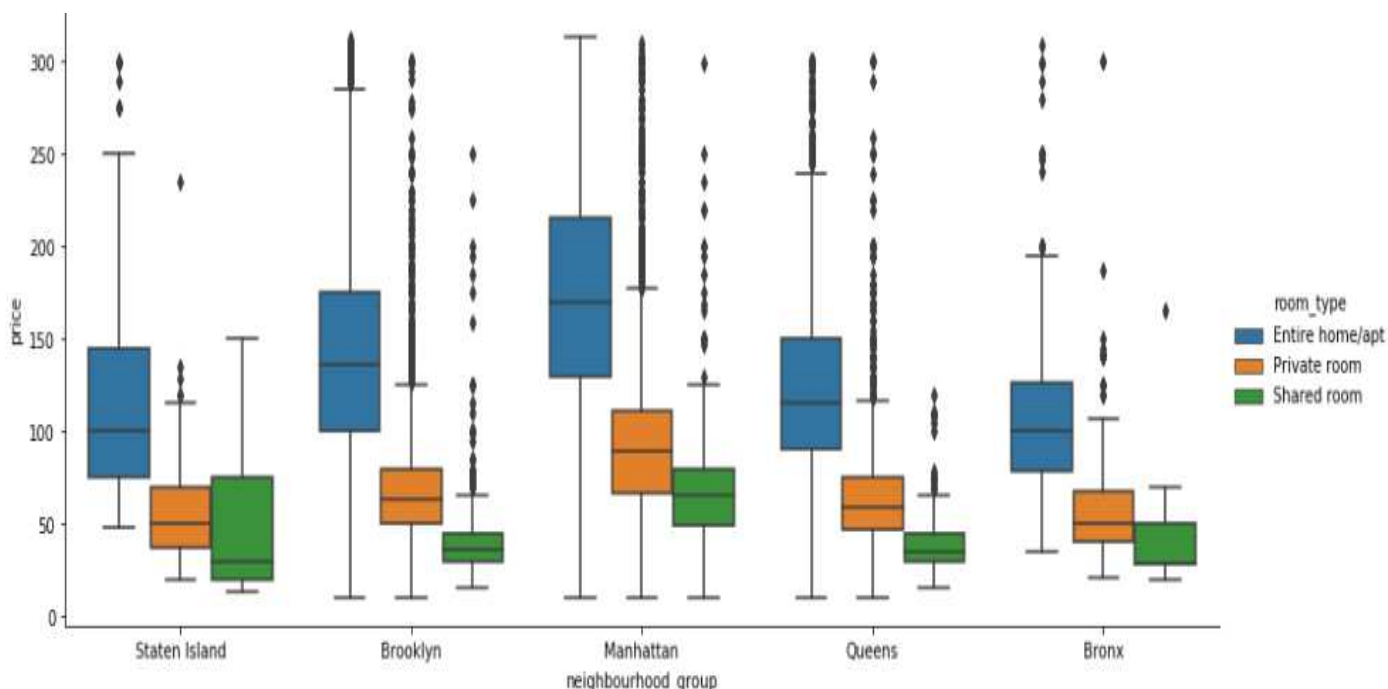
Al revisar la información total de la fuente de datos se toma en cuenta que la densidad de precios se presentar en el gráfico #1, el precio promedio ronda los \$167 +-\$64. Por cada ciudad del estado se presenta en la tabla#1:

	Cantidad_de_locales	Precio_Promedio	Precio_Minimo	Precio_Maximo	Precio_desv_estandar	Precio_25	Precio_50	Precio_75
neighbourhood_group								
Manhattan	15144	140.808637	10	313	66.307569	89	130.0	187
Brooklyn	15875	105.671181	10	312	59.121810	60	90.0	140
Queens	4507	87.590415	10	300	51.982260	50	70.0	105
Staten Island	312	87.163462	13	300	55.389897	50	73.5	105
Bronx	864	74.751157	20	309	44.516346	45	63.0	90

**Tabla #1: Resumen de los precios ofertados por ciudad en el Estado de Nueva York Airbnb 2019.**

La ciudad de Manhattan presenta los precios mayores con \$140 en promedio mientras que el estado de Bronx con un \$74.75 en promedio. Los Barrios con los precios más bajos son: New Dorp Beach , Little Neck , Bull's Head y Tremont con precios rondando los \$38 a \$50, y los más precios más altos son: Willowbrook, Neponsit, Breezy Point y Tribeca con precios desde \$194 a \$250.

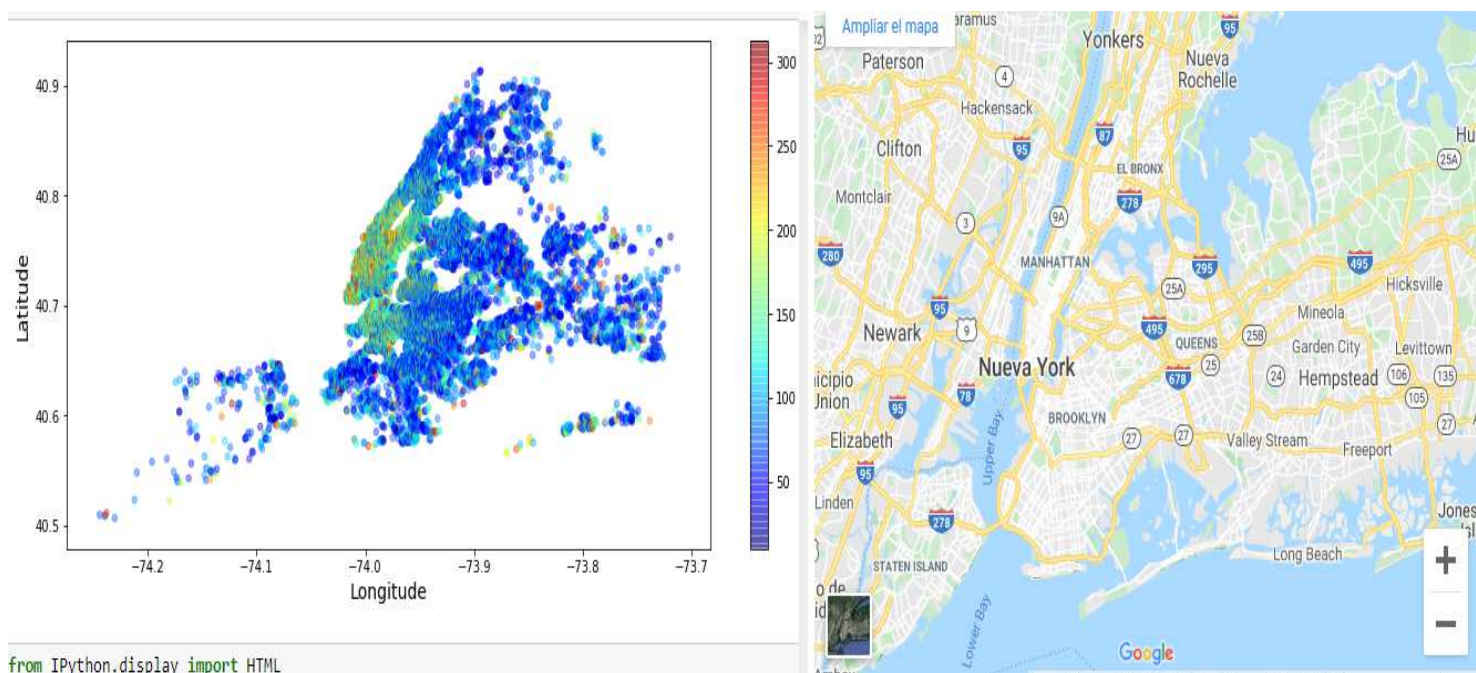
Cuando se toma en cuenta los tipos de cuarto: Compartido “shared”, Privado, “Private” y Apartamento Completo.



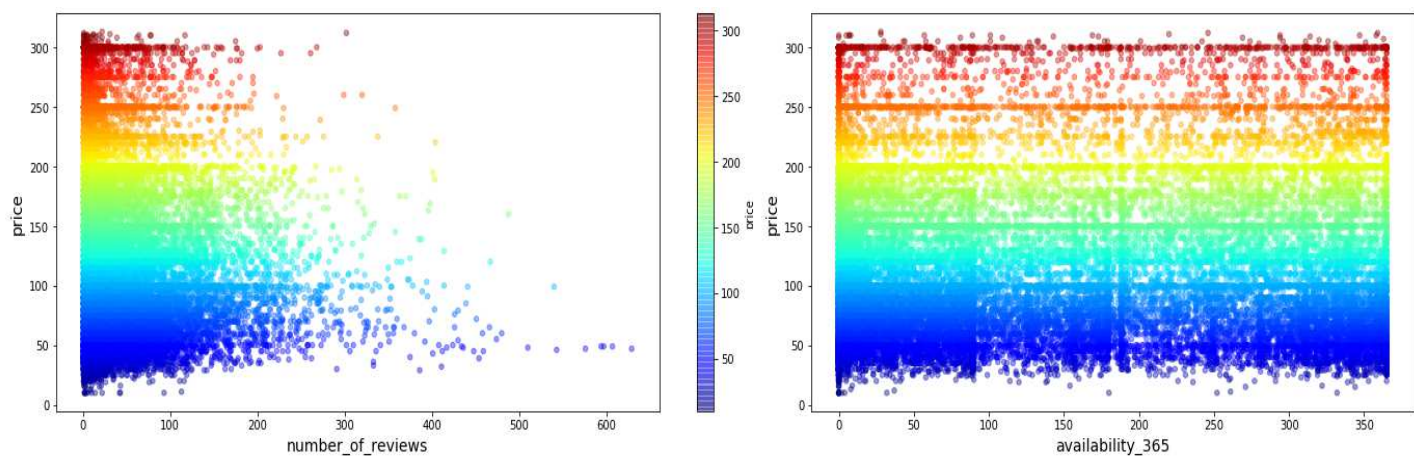
**Gráfico#2: Resumen de los precios ofertados por ciudad y tipo de cuarto en el Estado de Nueva York Airbnb 2019.**

Por Tipo de cuarto los precios más bajos son para los cuartos compartidos, en la ciudad de Bronx, Queens y Brooklyn con precios promedio de \$39,\$41 y \$44 respectivamente y los mas altos para el apartamento completo en las ciudades de Manhattan, Brooklyn y Queens con \$175,\$145 y \$129 respectivamente.

Al analizar por longitud y Latitud la distribución de precios se tiene lo mostrado en la fgura #1, la zona celeste denotan los precios de alrededor de \$150 en la zonas de Manhattan y Brooklyn. :



**Figura #1: Distribución espacial de precios.**



**Figura #2: Resumen características números de opiniones o “reviews” y disponibilidad de días durante el año “availability 365”.**

De las figuras detalladas, se nota que los atributos de longitud, latitud, número de revisiones y disponibilidad de días durante el años se muestran como buenos predictores para los precios más altos, donde suelen a ver menos precios altos conforme aumentan. En el caso de la disponibilidad se nota menos precios altos durante disponibilidad de todo el rango.

### **Modelos Probados.**

Para realizar el modelado de regresión lineal para predecir la variable “precio”, se procedió a codificar de manera adicional los atributos existentes, rescalando los atributos numéricos y se codificaron las categorías neighborhood\_group y room type con “one shot encoding”. Así los atributos finales para modelado y optimización son:

Atributos utilizados
latitude
longitude
minimum_nights
last_review_days
reviews_per_month
calculated_host_listings_count
availability_365
price
neighbourhood_group_Manhattan
neighbourhood_group_Queens
neighbourhood_group_Staten Island
room_type_Private room
room_type_Shared room

**Tabla #2 : Atributos utilizados para los modelos.**

Los resultados obtenidos de los modelos escogidos se presentan a continuación:

Modelos Utilizados	MSE	R2
GradientBoostingRegressor	1731,96	0,58
LinearRegression	2099,96	0,49
SVR	2150,27	0,48
LogisticRegression	2408,53	0,42
DecisionTreeClassifier	2718,97	0,34
KNN	2940,91	0,29

**Tabla #3: Resultados de los modelos utilizados con datos de prueba**

### **Resultados del Modelo Escogido.**

Una vez realizado los modelos, se procede a predecir los resultados con los datos de prueba. Comparando los valores predecidos con los reales.

Los modelos que se comportan mejor son los modelos de Gradient Boosting Regresor y Lineal Regresor debido al valor MAE y R2 de acuerdo a la tabla#3.

### **Conclusiones y Recomendaciones Adicionales.**

Como se podía observar en los gráficos y figuras anteriores existe una relación entre las variables de ciudad o grupo de barrios ("neighborhood group"), barrios, latitud longitud , tipo de cuarto, cantidad de opiniones o "reviews" y disponibilidad de días, que de acuerdo a los datos presentados permiten hacer estimaciones e inferencias importantes sobre el precio de la localidad.

Para obtener estimaciones más precisas se deben determinar si existen otros atributos que reflejen mejor la determinación del precio de la localidad. Amenidades internas, cantidad de cuartos, servicios adicionales por ejemplo. Esto debido a que con los atributos a manos se logró una valor máximo de 1731,96 de MAE y 0,58 de R2 con el algoritmos Gradient Boost.

En futuras iteraciones se puede analizar el impacto que tienen los dueños con más de una localidad de alquiler, verificando como se comportan los precios de este tipo de dueños con respecto a los que tiene una sola localidad.