

UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

INTRODUCCIÓN A LA CIENCIA DE DATOS

INFORME EDA

AIRBNB PROYECTO

ANGEL HURTADO VALDEZ

RODRIGO AMAYA

LUCIA CARTAGENA MIRANDA

OSWALDO QUISPE

SERGIO RICCE

Número de palabras:

DS1000

ÍNDICE

1. INTRODUCCIÓN

2. ESCENARIO

3. MÉTODOS

3.1. Limpieza de datos

3.2. Gráficos

3.3. No gráficos

4. CONCLUSIONES

5. BIBLIOGRAFÍA

1. INTRODUCCIÓN

Airbnb es una compañía que mediante su plataforma de alojamientos virtual emplea la economía colaborativa, un modelo económico empleando la tecnología para prestar, comprar, vender o compartir o alquilar bienes servicios, una alternativa a los hoteles tradicionales. Por su parte, los anfitriones pueden publicitarse empleando sus propiedades para los huéspedes, una característica importante y característica, los huéspedes y anfitriones pueden valorarse mutuamente, considerando que tan buen anfitrión has sido, las condiciones de la habitación, el aseo de la misma; mientras que, califican si has sido un huésped que cuida de los ambientes, tranquilo y si has cumplido con las reglas del alojamiento, entre otros.

En este caso, nuestra población de estudio será Belice, un país de Centroamérica que cuenta con una población de 400 031 personas en 2021, con una actividad turística alta, debido a la diversidad de su fauna marina.

En este caso, recopilaremos datos de Airbnb, empleando datos actualizados hasta el 23 de septiembre del 2023. Por ello, realizaremos un estudio de los alojamientos de la ciudad de Béliçe, porque su población es reducida y posee bastante información acerca de alojamientos. Por ende, realizaremos un estudio empleando las variables de precio, número de reseñas, media de precios, tiempo, zona, tipo de habitación, número de estrellas, entre otros.

El objetivo principal de este informe es encontrar la relación entre los datos es identificar un patrón entre el precio y la locación con respecto al resto de variable, de esta forma encontrar si existe una tendencia entre los usuarios y dar a conocer información valiosa para que los próximos alojamientos particulares sean más eficientes al salir en la plataforma.

2. ESCENARIO

Belice es un país con un gran atractivo turístico, debido a su diversidad de especies marinas y su fauna silvestre, por ello, muchos de los turistas buscan alojarse en zonas específicas o dónde haya mayor cercanía a estos lugares turísticos. Por ello, realizaremos un estudio de los patrones de los turistas en las distintas regiones de Belice, en base a la tendencia con respecto a variables de precio, calificación, región de alojamiento, entre otros, consideraremos las zonas más turísticas y que más atraen a los turistas.

En síntesis, la investigación es aplicable a empresas o individuos interesados en lo turístico que indagan en las tendencias de los turistas, el propósito es facilitar la selección de un área para arrendamiento o considerar alquilar su hogar basado en esta tendencia para

garantizar su éxito.

3. MÉTODOS

3.1. Limpieza de datos

Figura N° 1

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3100 entries, 0 to 3099
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    3100 non-null   int64
1   price                 3100 non-null   int64
2   number_of_reviews     3100 non-null   int64
3   last_review           2330 non-null   object
4   neighbourhood          3100 non-null   object
5   name                  3100 non-null   object
6   availability_365       3100 non-null   int64
7   host_id               3100 non-null   int64
8   room_type             3100 non-null   object
9   latitude              3100 non-null   float64
10  longitude             3100 non-null   float64
dtypes: float64(2), int64(5), object(4)
memory usage: 266.5+ KB
```

Elaboración propia

Para este proyecto, cargamos listings.csv que tiene 18 columnas. Solo vamos a utilizar 11 de sus columnas para progresar. Se verifica que cada columna tenga su tipo adecuado.

Figura N° 2

	id	price	number_of_reviews	last_review	neighbourhood	name	availability_365	host_id	room_type	latitude	longitude
3095	979088845313929626	66	0	NaN	Toledo Mainland	Cabin in Ymeri Grove - ★New - 1 bedroom - 2 b...	269	150789	Private room	16.161079	-88.874180
3096	979657301560948512	1686	0	NaN	Belize Islands	Cabin in San Pedro - ★New - Studio - 2 baths	7	55997332	Private room	17.921394	-87.960042
3097	980639968759012759	186	0	NaN	Cayo	Home in Spanish Lookout - ★New - 2 bedrooms - ...	269	234363583	Entire home/apt	17.243793	-89.029992
3098	983265546231090140	353	0	NaN	Cayo	Home in Calla Creek - ★New - 1 bedroom - 2 bed...	364	408053114	Entire home/apt	17.155715	-89.133064
3099	984345079968281304	121	0	NaN	Cayo	Home in San Ignacio - ★New - 3 bedrooms - 4 be...	270	462323256	Entire home/apt	17.154598	-89.078015

Elaboración propia

Acá comprobamos el país de las propiedades. En la función En País, veremos si el país de cada propiedad es Belice.

Figura N° 3

	id	price	number_of_reviews	last_review	neighbourhood	name	availability_365	host_id	room_type	latitude	longitude	coords	pais
0	36690	91	409	2023-09-20	Belize Islands	Tiny home in Caye Caulker ★4.82 · 1 bed · 1 ...	165	157752	Entire home/apt	17.74832	-88.02474	17.74832,-88.02474	Belize
1	120115	603	56	2023-05-20	Belize Islands	Condo in Caye Caulker Village ★4.85 · 2 bed...	235	604999	Entire home/apt	17.74166	-88.02395	17.74166,-88.02395	Belize

Elaboración propia

A cada propiedad, se le asignan coordenadas (coords) que son la latitud y longitud unidos en un string, y con estas coordenadas, se obtiene su país. Y La verificación del país de las propiedades dura más de 30 minutos, por lo que el df obtenido anteriormente, fue convertido a csv con el fin de usar los datos de forma más rápida en el futuro.

Figura N° 4

```
Belize      3099
temala      1
Name: pais, dtype: int64
```

Elaboración propia

Leemos el nuevo csv y contamos las propiedades.

Figura N° 5

```
3099
```

Elaboración propia

Se observan valores distintos a Belize con el .value_counts(). Solo buscamos contar los que son de Belize. Serían un total de 3099 propiedades.

Figura N° 6

```
id          0
price       0
number_of_reviews  0
last_review 770
neighbourhood  0
name        0
availability_365  0
host_id     0
room_type   0
latitude    0
longitude   0
coords      0
pais        0
dtype: int64
```

Elaboración propia

Existen 770 datos nulos en last_review, lo que significa que sus establecimientos no recibieron reseñas.

Figura N° 7

```
id          0
price       0
number_of_reviews  0
last_review  0
neighbourhood  0
name        0
availability_365  0
host_id     0
room_type   0
latitude    0
longitude   0
coords      0
pais        0
dtype: int64
Cantidad de datos: 2321
```

Elaboración propia

Se decidió eliminar los valores cuyo precio sea más de 3000 dólares beliceños y cuyas

reseñas sean 0. Se puede observar como ya no hay valores nulos en last_review. Los valores ahora disponibles son 2321.

Figura N° 8

	id	price	number_of_reviews	last_review	neighbourhood	name	availability_365	host_id	room_type	latitude	longitude
0	36690	91	409	2023-09-20	Belize Islands	Tiny home in Caye Caulker ★4.82 · 1 bed · 1 ...	165	157752	Entire home/apt	17.74832	-88.02474

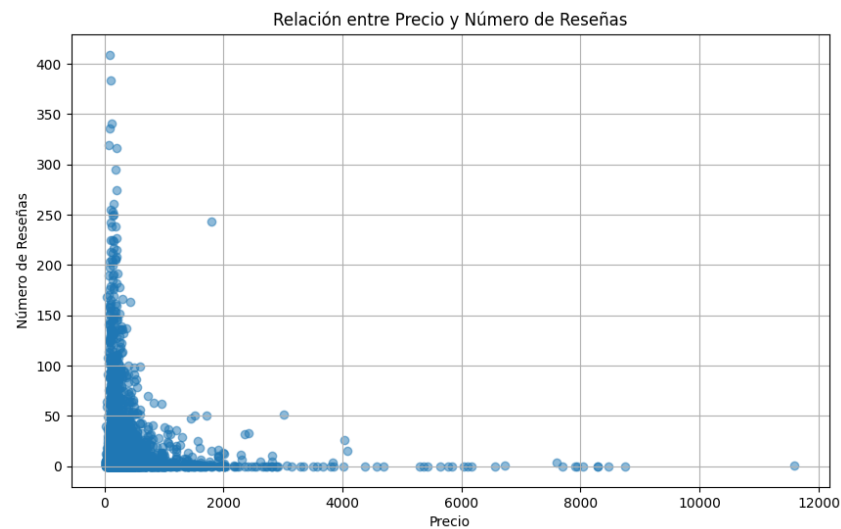
Elaboración propia

Una vez tenemos los datos limpios, quitamos las columnas que no se utilizarán (coords y pais).

3.2. Gráficos

Figura N° 9

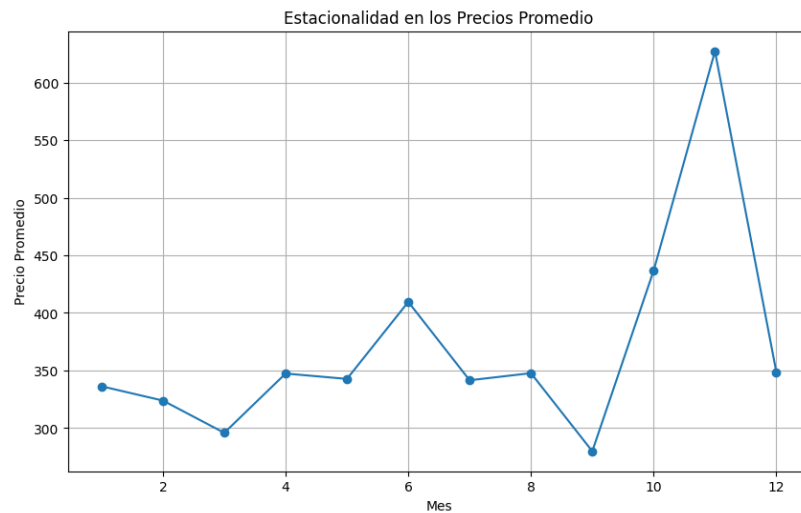
Gráfico de Dispersión entre Precio y Número de Reseñas



Elaboración propia

Los precios mientras más bajos sean, más demandas tendrán y por lo tanto habrá más reseñas, y eso se demuestra desde el primer punto donde vemos que si el precio es bajo hay bastantes reseñas. En otras es una correlación negativa

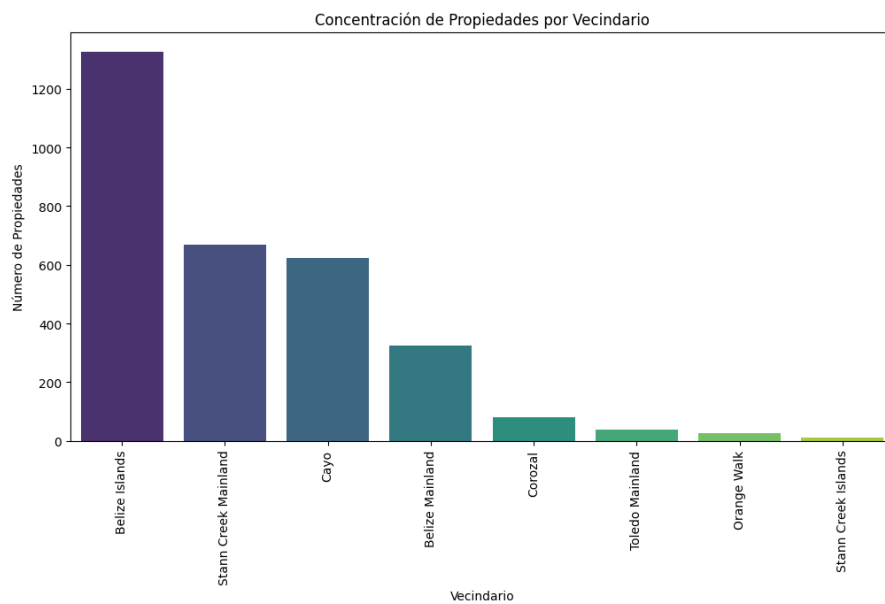
Figura N° 10
Estacionalidad en los Precios Promedio



Elaboración propia

Dependiendo del mes, los precios promedio suben considerablemente, por ejemplo, las personas suelen viajar en octubre o noviembre por vacaciones de trabajo o estudio, o por ofertas; a su vez, las festividades siempre suelen venir con un montón de ofertas.

Figura N° 11
Gráfico de Barras Agrupados entre Vecindario y Número de Propiedades

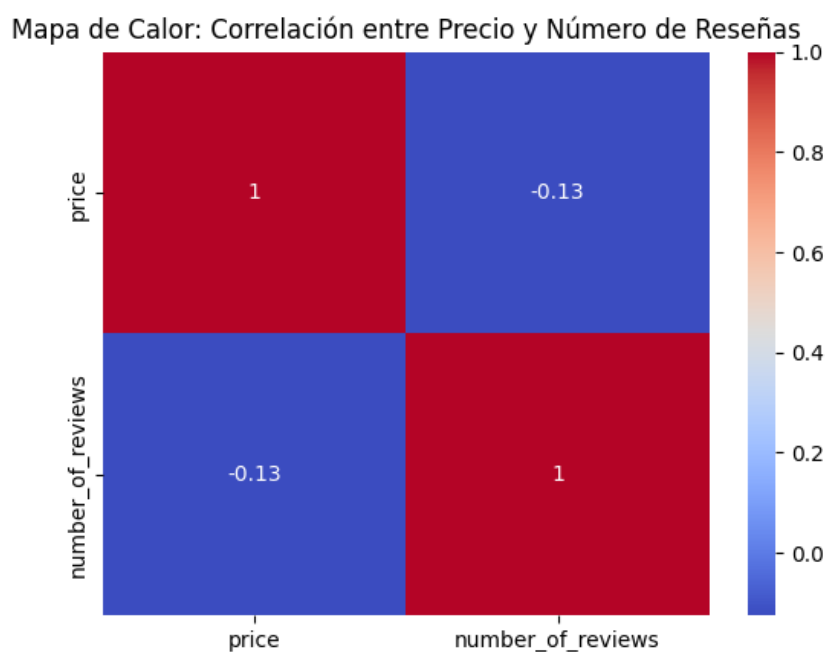


Elaboración propia

En ciertas zonas de Belice son demandadas por el espacio y el atractivo de este, aquellas zonas también son decisivas por el lugar donde se encuentra, ya que los clientes siempre buscan lugares accesibles o prácticos.

Figura N° 12

Mapa de calor: Correlación entre Precio y Número de reseñas

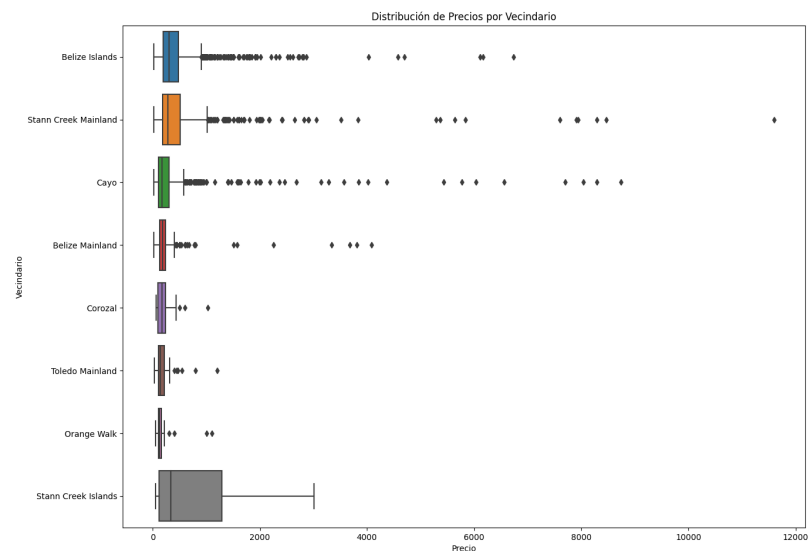


Elaboración propia

El mapa de calor identifica que existe una correlación negativa débil en los datos de precio y número de reseñas, es decir, a menor sea el precio, existe una mayor cantidad de reseñas, básicamente los datos son inversamente proporcionales.

Figura N° 13

Gráfico Side-by-Side Boxplots entre Belize Islands, Stann Creek Mainland, Cayo, Belize Mainland, Corozal, Toledo Mainland, Orange Walk, Stann Creek Islands.



Elaboración propia

Dependiendo del vecindario, el precio puede subir o bajar, y eso normalmente está fijado por la calidad del lugar y el cómo son tratados ahí, a su vez también está definido por el lugar por donde se encuentra.

3.3. No gráficos

Figura N° 14

Precio promedio por tipo de habitación

Precio Promedio: 423.75870967741935
Precio Mediana: 246.0

	room_type	price
0	Entire home/apt	484.283349
1	Hotel room	238.773109
2	Private room	308.572110
3	Shared room	91.821429

Elaboración propia

Figura N° 15

DataFrame por el número de reseñas en orden descendente

	name	number_of_reviews
0	Tiny home in Caye Caulker · ★4.82 · 1 bed · 1 ...	409
305	Cottage in San Ignacio · ★4.95 · 1 bedroom · 1...	383
462	Rental unit in Belize City · ★4.89 · 1 bedroom...	341
332	Cabin in San Ignacio · ★4.94 · 1 bedroom · 1 b...	336
470	Cabin in San Ignacio · ★4.92 · 2 bedrooms · 2 ...	319
167	Home in Ladyville · ★4.91 · 2 bedrooms · 2 bed...	316
230	Home in Caye Caulker · ★4.92 · 1 bedroom · 1 b...	295
42	Guesthouse in San Pedro · ★4.82 · 1 bedroom · ...	274
40	Rental unit in Caye Caulker · ★4.96 · Studio ...	261
175	Cabin in Belize District · ★4.80 · 1 bedroom ...	255

Elaboración propia

Figura N° 16

Los datos por vecindario y la disponibilidad promedio

	neighbourhood	availability_365
0	Belize Islands	264.617647
1	Belize Mainland	261.987692
2	Cayo	233.477492
3	Corozal	262.375000
4	Orange Walk	277.962963
5	Stann Creek Islands	236.666667
6	Stann Creek Mainland	252.710015
7	Toledo Mainland	174.076923

Elaboración propia

Figura N° 17

Los anfitriones por el número de propiedades en orden descendente

	Host ID	Número de Propiedades
0	76414486	231
1	261868426	62
2	151270790	36
3	496944100	33
4	604999	30
5	1120393	29
6	152891603	27
7	144130330	26
8	88200158	23
9	85339670	22

Elaboración propia

Figura N° 18

Correlación entre el precio y el número de reseñas y covarianza entre el precio y el número de reseñas

Correlación: -0.1257429511419175
Covarianza: -3543.924219467258

Elaboración propia

Figura N° 19

Tabla de tabulación cruzada

neighbourhood	Belize Islands	Belize Mainland	Cayo	Corozal	Orange Walk	\
room_type						
Entire home/apt	990	168	386	53	17	
Hotel room	46	22	27	1	0	
Private room	282	131	201	26	10	
Shared room	8	4	8	0	0	

neighbourhood	Stann Creek Islands	Stann Creek Mainland	Toledo Mainland
room_type			
Entire home/apt	2	476	22
Hotel room	0	23	0
Private room	10	162	17
Shared room	0	8	0

Elaboración propia

4. CONCLUSIONES

Entre room_type y price.-

Entre los valores "room_type" y "price" se infiere que los precios promedio varían según el tipo de habitación disponible. Por ejemplo, para "Entire home/apt" el precio promedio es aproximadamente 364.85; a comparación de "Shared room" que el precio promedio es aproximadamente 58.91, por lo tanto uno tiende a ser más caro que el otro. Esta información nos es útil para hacer una proyección de lo que tendremos que pagar dependiendo del tipo de cuarto que queramos

Entre name y number_of_reviews.-

Aquí identificamos 10 de los hospedajes mejor rankeados según la cantidad de reseñas que han recibido. Lo cual nos llegaría a ser muy útil para comprender cuáles son los hospedajes más revisados, y por lo tanto, posiblemente los más populares.

Entre neighbourhood y availability_365.-

Nos proporciona información sobre la disponibilidad promedio durante el año de alojamientos en diferentes áreas geográficas. Y eso nos puede ser útil para comprender cómo es que varía la disponibilidad de los distintos hospedajes en relación con el área geográfica y

tomar una decisión basada en esta información.

Entre Host ID y Número de Propiedades.-

Nos muestra los 10 primeros Host con la mayor cantidad de número de propiedades a su disposición.

Entre la Correlación y la Covarianza.-

Hacemos una correlación y covarianza entre el precio y la cantidad de reseñas.

La correlación es de aproximadamente -0.1632, esto quiere decir que es débilmente negativa, esto nos indica que, en promedio, los hospedajes más caros tienen una menor cantidad de reseñas y viceversa.

La covarianza es de aproximadamente -2189.31, esto quiere decir que se mueven en sentido inverso, lo cual nos refuerza la idea de que si hay una relación negativa entre el precio y la cantidad de reseñas.

Acerca de la tabla cruzada.-

Entre neighbourhood y room_type podemos interpretar que dependiendo del tipo de cuarto que busquemos, habrá una mayor o menor cantidad de hospedajes disponibles en las distintas áreas geográficas que queremos buscar.

En conclusión, los resultados del análisis de datos concluyen que los factores de precio y locación son realmente influyentes con respecto al resto de variables, puesto que, afectan en gran medida las decisiones del consumidor. Esto se evidencia mediante el web scraping, dando como resultado que el precio tiene una correlación ligeramente negativa con respecto al número de reseñas, a menor sea el precio, más reseñas tiene; asimismo, la variabilidad del precio con respecto al mes del año. Mientras que, la locación se encuentra relacionada con el número de alojamientos, puesto que, mientras más comercial sea una zona resulta más rentable tener propiedades en esa zona; por otro lado, la locación también se relaciona con la variabilidad del precio en un alojamiento, considerando que, a más solicitado sea un lugar, más está dispuesto a pagar el consumidor. En síntesis, el web scraping resulta de gran utilidad para realizar un análisis exploratorio del mercado y la información que nos proporciona, de esta forma se puede realizar un análisis y destacar algunos resultados, como la gran influencia de la variable precio y locación en la opinión de los consumidores.

