

**Estimación de Precios Competitivos
para Propiedades en AirBnB CDMX**
Deep Learning



Equipo 7

Raúl Gerardo Reyes Barrón 192129

César Armando Rojas Flores 220019

León Felipe Gómez Zarza 111203

Ulises Reyes García 152113

Emanuel Ortiz Bassoco 130669

Introducción

Planteamiento del problema

¿Quién es **nuestro cliente**?



Grupo de propietarios de inmuebles en CDMX registrados en AirBnB

¿Qué está **buscando**?



Establecer precios competitivos con base en la zona y características del inmueble

¿Qué **retos** enfrenta?



Realiza análisis manual y carece de información suficiente para establecer una estrategia de precios adecuada y efectiva

Introducción



Contexto del Modelo Anterior

Variables Utilizadas		
✓ 69 variables		
31 Numéricas		
39 Categóricas		
1 Objetivo		
dist_to_Zocalo	bedrooms	room_type_Hotel_room
dist_to_Angel_de_la_Independencia	beds	room_type_Shared_room
dist_to_Parque_México	amenities_n	bathrooms_type_unespecified
dist_to_Bosque_de_Chapultepec	calculated_host_listings_count_entire_homes	bathrooms_type_private
dist_to_Coyoacan_Centro	calculated_host_listings_count_private_rooms	bathrooms_type_shared
dist_to_Aeropuerto_Internacional	calculated_host_listings_count_shared_rooms	alcaldia_Cuauhtémoc
dist_to_Monumento_a_la_Revolucion	accommodates	alcaldia_Cuajimalpa_de_Morelos
dist_to_Museo_Soumaya	description	alcaldia_Miguel_Hidalgo
dist_to_Santa_Fe	neighborhood_overview	alcaldia_Coyoacán
dist_to_Estadio_Azteca	host_location	alcaldia_Alvaro_Obregon
dist_to_UNAM	host_about	alcaldia_Benito_Juárez
dist_to_Centro_Comercial_Perisur	host_is_superhost	alcaldia_Iztacalco
dist_to_Palacio_de_Bellas_Artes	instant_bookable	alcaldia_Tlalpan
dist_to_Auditorio_Nacional	host_verifications_email	alcaldia_La_Magdalena_Contreras
dist_to_Basilica_de_Guadalupe	host_verifications_phone	alcaldia_Iztapalapa
host_since_days	host_verifications_work_email	alcaldia_Venustiano_Carranza
host_since_days	host_verifications_no_verifications	alcaldia_Gustavo_A._Madero
host_since_months	host_response_time_within_an_hour	alcaldia_Xochimilco
bathrooms_numeric	host_response_time_a_few_days_or_more	alcaldia_Azcapotzalco
host_total_listings_count	host_response_time_within_a_day	alcaldia_Tláhuac
host_acceptance_rate	host_response_time_within_a_few_hours	alcaldia_Milpa_Alta
host_response_rate	host_response_time_no_response_time	price
latitude	room_type_Entire_home/apt	
longitude	room_type_Private_room	

Alcance

Sólo para propiedades nuevas (sin listarse previamente en AirBnB)

No se procesó texto (se trataron como categóricas o se omitieron)

Desempeño

Modelo Ganador:
Random Forest

MAPE:
25%

Introducción

Objetivos



OBJETIVO PRINCIPAL

- ✓ **Estimar el precio de renta** por noche en AirBnB del inmueble de cada cliente con base en su ubicación y características generales



OBJETIVOS SECUNDARIOS

- ✓ Utilizar el modelo inicial como base para implementar una nueva capa de mejoras y mejorar el MAPE obtenido previamente (25%)
- ✓ Utilizar como variables explicativas las reseñas (en score y texto) y el contexto de las amenidades
- ✓ Incorporar en el alcance tanto propiedades nuevas en AirBnB como también previamente listadas (así como sus respectivos indicadores)

Introducción

Alcance y limitaciones

Alcance



Análisis se restringirá únicamente a propiedades en CDMX



Se utilizarán los mismos datos con los que se entrenó la 1ra versión de este modelo



Se utilizarán datos numéricos, categóricos y de texto que se consideren relevantes para el modelo

Limitaciones



Se propondrá un precio de renta diario fijo (sin estacionalidad o variabilidad)

Modelado

Datos y herramientas

Base de datos de propiedades

Archivos en formato
.CSV

26,536 registros
(propiedades)

Tabla de:
26,536 x 75

75 posibles variables



Base de datos de reseñas

Longitud Promedio de
cada review: 190
caracteres

Número de reviews:
1,152,978

Número de palabras:
224,286,344

Modelado



Listado de variables original

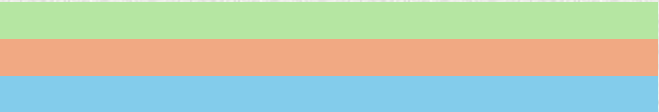
neighbourhood_group_cleansed	id	listing_url
latitude	scrape_id	last_scraped
longitude	host_id	source
bathrooms	host_listings_count	name
bedrooms	host_total_listings_count	description
beds	accommodates	neighborhood_overview
minimum_nights_avg_ntm	minimum_nights	picture_url
maximum_nights_avg_ntm	maximum_nights	host_url
calendar_updated	minimum_minimum_nights	host_name
review_scores_rating	maximum_minimum_nights	host_since
review_scores_accuracy	minimum_maximum_nights	host_location
review_scores_cleanliness	maximum_maximum_nights	host_about
review_scores_checkin	availability_30	host_response_time
review_scores_communication	availability_60	host_response_rate
review_scores_location	availability_90	host_acceptance_rate
review_scores_value	availability_365	host_is_superhost
license	price	host_thumbnail_url
reviews_per_month	has_availability	host_picture_url
number_of_reviews_l30d	host_neighbourhood	neighbourhood_cleansed
calculated_host_listings_count	host_verifications	property_type
calculated_host_listings_count_entire_homes	host_has_profile_pic	room_type
calculated_host_listings_count_private_rooms	host_identity_verified	bathrooms_text
calculated_host_listings_count_shared_rooms	neighbourhood	amenities
number_of_reviews	calendar_last_scraped	last_review
number_of_reviews_ltm	first_review	instant_bookable

✓ 34 object

✓ 23 integer

✓ 18 float

float64
int64
object

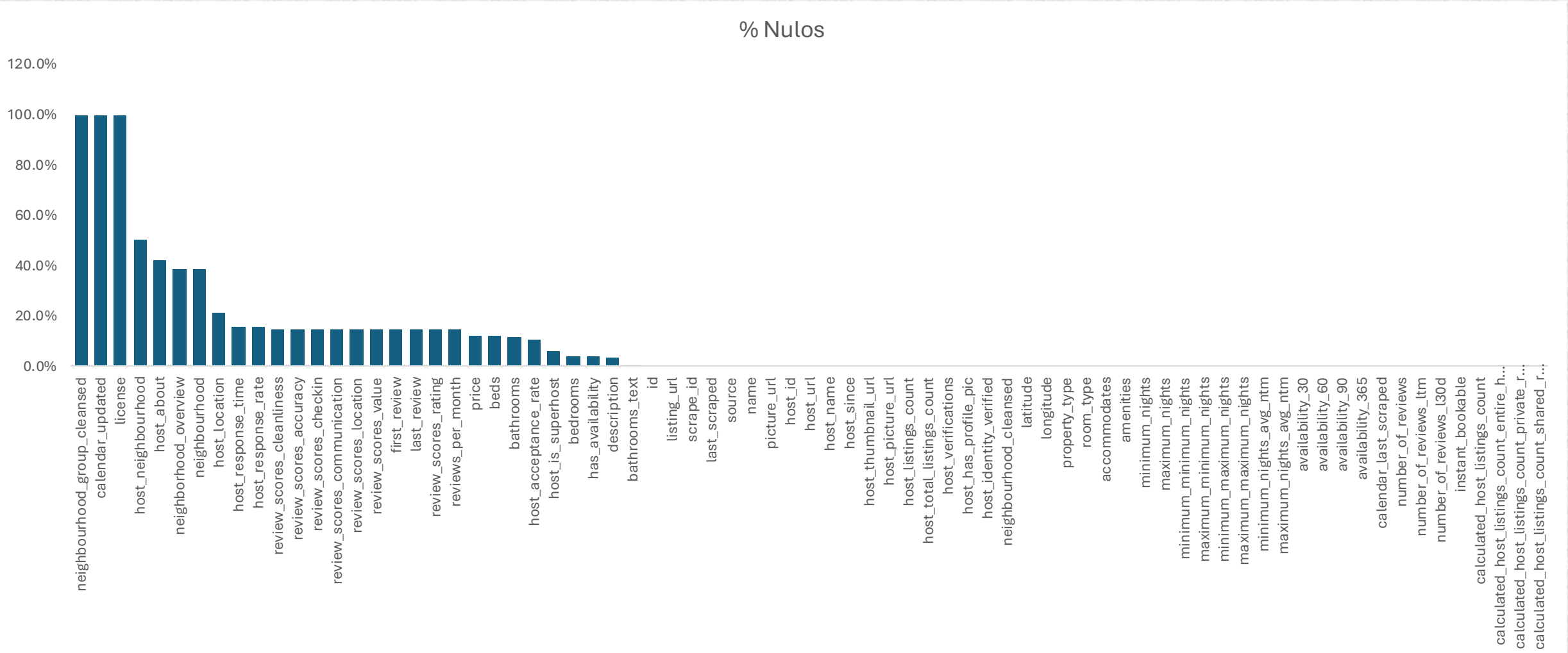


Resumen estadístico (numéricos y categóricos)

Métricas	host_listings_count	host_total_listings_count	latitude	longitude	accommodates	bathrooms	bedrooms	beds	minimum_nights	maximum_nights	minimum_minimum_nights	maximum_minimum_nights	minimum_maximum_nights	maximum_maximum_nights	minimum_nights_avg_ntm	maximum_nights_avg_ntm	availability_30	availability_60	availability_90
mean	26.7	36.7	19.4	-99.2	3.3	1.5	1.6	1.9	4.7	546.6	4.2	4.8	752.1	772.3	4.4	763.1	15.8	36.7	59.2
std	83.1	150.6	0.0	0.0	2.3	1.0	1.3	1.7	24.2	431.0	23.3	25.5	461.7	468.4	23.4	468.5	11.6	22.0	31.8
min	1.0	1.0	19.2	-99.3	1.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
25%	1.0	2.0	19.4	-99.2	2.0	1.0	1.0	1.0	1.0	190.0	1.0	1.0	365.0	365.0	1.0	365.0	2.0	20.0	39.0
50%	4.0	5.0	19.4	-99.2	2.0	1.0	1.0	1.0	2.0	365.0	1.0	2.0	1,125.0	1,125.0	2.0	1,125.0	18.0	45.0	72.0
75%	14.0	17.0	19.4	-99.2	4.0	2.0	2.0	2.0	2.0	1,125.0	2.0	3.0	1,125.0	1,125.0	2.4	1,125.0	27.0	57.0	87.0
max	1,776.0	9,003.0	19.6	-99.0	16.0	49.0	50.0	50.0	1,125.0	1,825.0	1,125.0	1,500.0	9,999.0	9,999.0	1,125.0	9,999.0	30.0	60.0	90.0

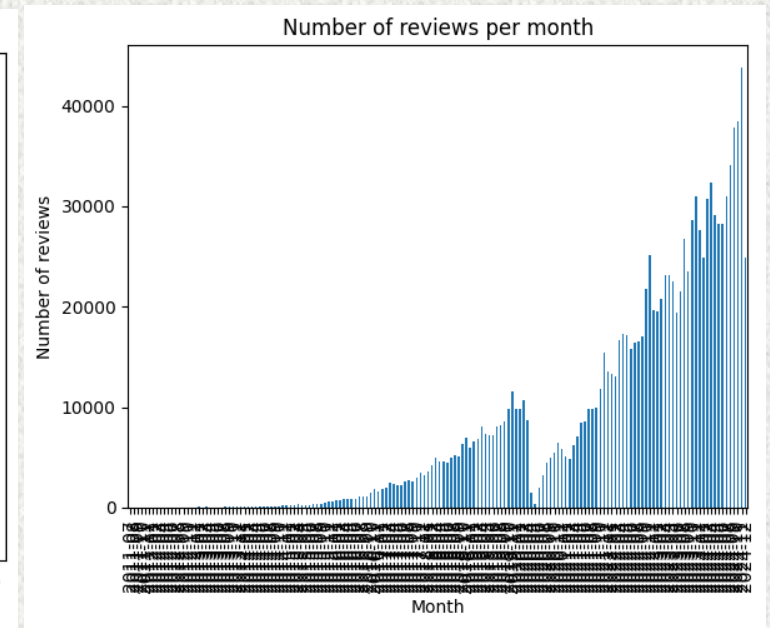
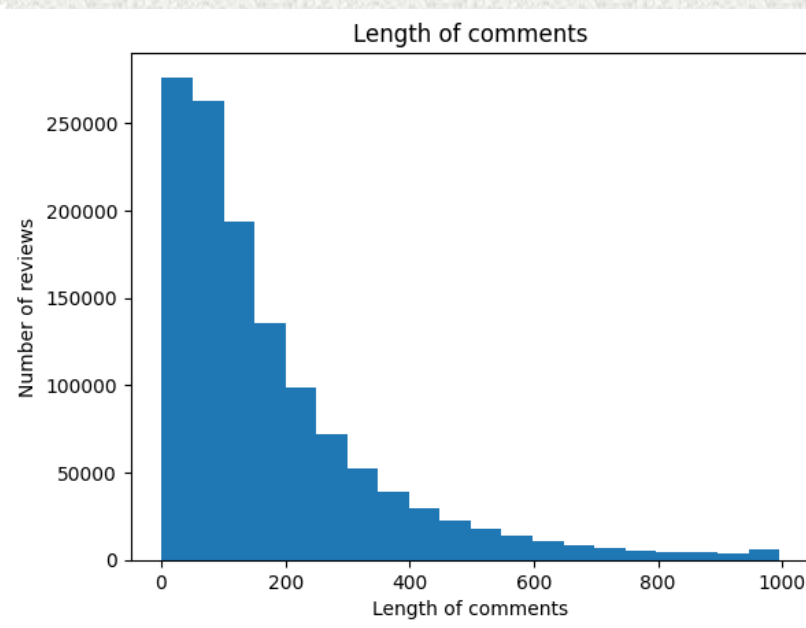
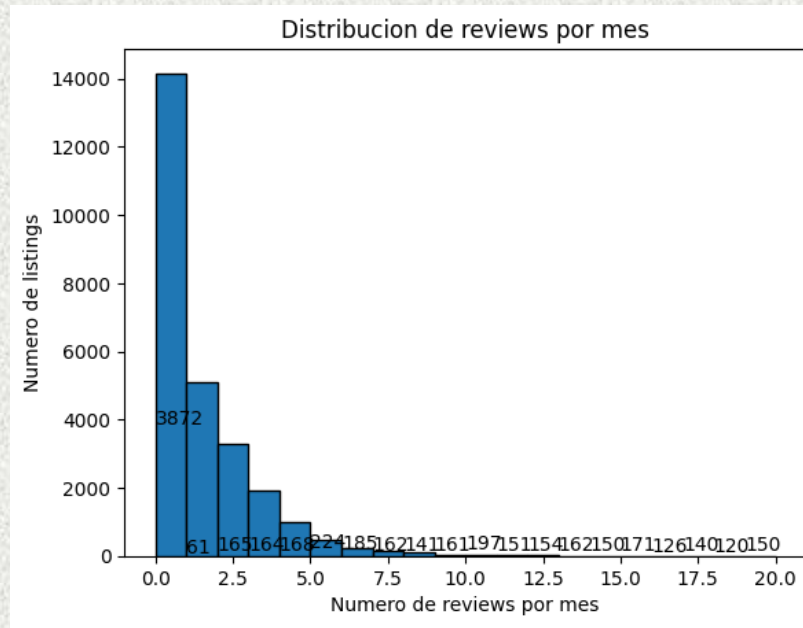
Métricas	availability_365	number_of_reviews	number_of_reviews_ltm	number_of_reviews_l30d	review_scores_rating	review_scores_accuracy	review_scores_cleanliness	review_scores_checkin	review_scores_communication	review_scores_location	review_scores_value	calculated_host_listings_count	calculated_host_listings_count_entire_homes	calculated_host_listings_count_private_rooms	calculated_host_listings_count_shared_rooms	reviews_per_month
mean	224.2	43.4	13.6	0.9	4.8	4.8	4.8	4.8	4.8	4.8	4.7	17.2	14.2	2.5	0.4	1.7
std	124.7	73.2	19.9	1.7	0.4	0.3	0.4	0.3	0.3	0.3	0.4	43.3	42.2	6.9	3.7	1.9
min	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
25%	116.0	3.0	0.0	0.0	4.7	4.8	4.7	4.8	4.8	4.8	4.7	1.0	0.0	0.0	0.0	0.4
50%	261.0	15.0	5.0	0.0	4.9	4.9	4.9	4.9	4.9	4.9	4.8	3.0	1.0	0.0	0.0	1.1
75%	345.0	53.0	20.0	1.0	5.0	5.0	5.0	5.0	5.0	5.0	4.9	11.0	6.0	2.0	0.0	2.4
max	365.0	1,277.0	382.0	53.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	267.0	267.0	61.0	53.0	44.4

Características de los datos (numéricos y categóricos)



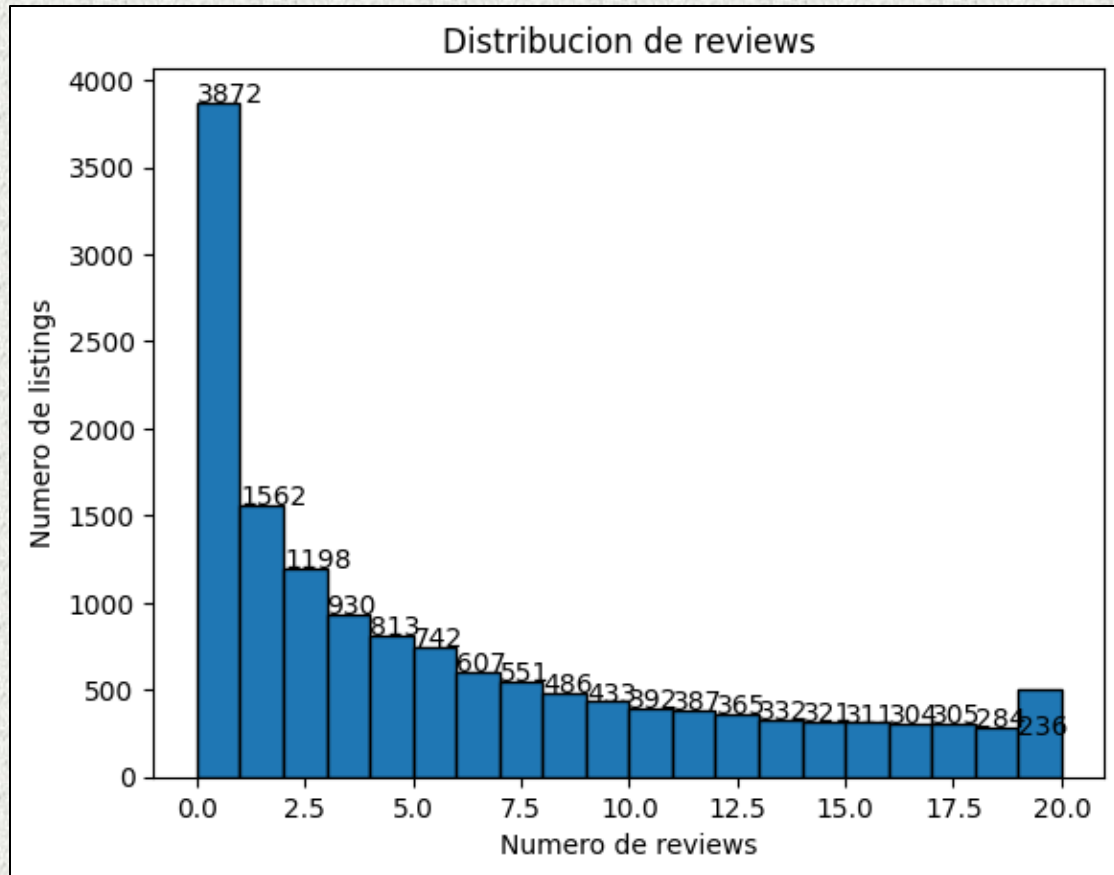
Modelado

Características de los datos (reseñas en texto)

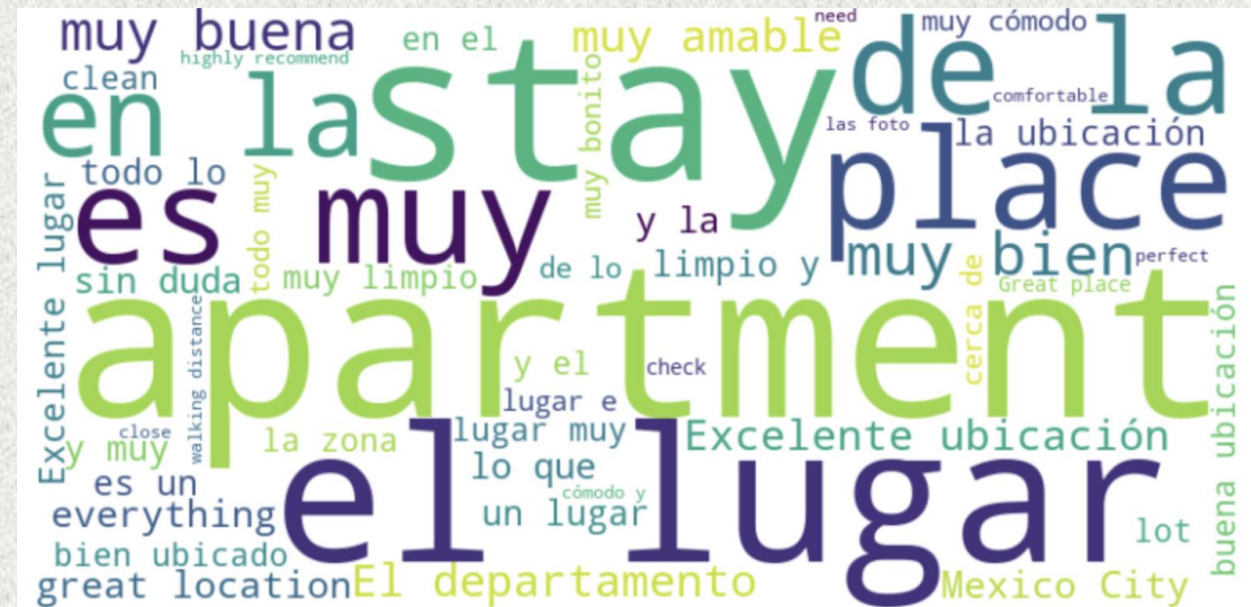


Modelado

Características de los datos (reseñas en texto)



- 1,152,978 reviews con un total de 224,286,344 palabras.



Modelado

Solución propuesta

Tenemos 3 tipos de información

Características de la Propiedad

Variables
Numéricas

Estandarizar

Variables
Categorías

One-Hot

Capas Ocultas Densas (ReLU*)

Vector Salida
(len=32*)

Reseñas

Scores

Promediar

Estandarizar

Texto

Embedding
(BERT*)

LSTM*

Vector Salida
(len=64*)

Listado de Amenidades

Limpiar y lematizar

Crear 10 categorías

Agrupar semánticamente

Count Encoder

Capas Ocultas Densas (ReLU*)

Vector Salida
(len=16*)

Red Neuronal Multi-Input

Precio del
AirBnB

*Valores y herramientas sólo ejemplificativos