




AIRBNB-USA

PRESENTACION DATA SCIENCE -CODERHOUSE



Los sets de datos desarrollados  
aquí fueron obtenidos en kaggle  
sobre las rentas de inmuebles en  
USA de una empresa en auge

---

***CODERHOUSE***

# TABLA DE CONTENIDOS :

- DESCRIPCION DEL NEGOCIO A ESTUDIAR
- OBJETIVOS DEL MODELO
- DESCRIPCION DE LOS DATOS
- ALGORITMO ELEGIDO
- METRICAS DE DESEMPEÑO
- OPTIMIZACION
- CONCLUSION FINAL

The logo for Coderhouse, featuring the word "CODERHOUSE" in a bold, italicized, sans-serif font. The text is white and is set against a black rectangular background. The entire logo is positioned on the right side of the slide, overlapping a large, white, brush-stroke-like shape that has a rough, torn edge.



# DESCRIPCION DEL NEGOCIO A ESTUDIAR

- **LOS DATASETS DE AIRBNB ABARCA UNA GRAN RED DE ALQUILERES A NIVEL EUROPEO Y AMERICANO , ES UNA EMPRESA QUE SURGIO PARA VIAJANTES O EXTRANJEROS QUE NO QUERIA NRENTAR GRANDES LUGARES O LUGARES IBLIGADOS POR MAS DE 1 SEMANA.**
- **ESTA EMPRESA PERMITE RENTAR UN CUARTO O HABITACION POR HORAS ,MINUTOS O DIAS , MUCHAS DE LAS PERSONAS QUE LOS ARRIENDAN SON PERSONAS QUE NO LOS USAN.**
- **TRABAJAN CON EL REAL STATE PARA LUGARES MAS ESPECIFICOS O GENTE CON MAYORES REQUISITOS**
- **LO QUE TRATAREMOS DE PREDECIR ES QUE CONVIENE MAS A LA HORA DE RENTAR YA SEA SI USTED ES VIAJANTE , PERMANENTE O NO Y QUE LUGARES SON MAS ACCESIBLES AL BOLSILLO DE ACUERDO AL BARRIO ZONA O CIUDAD.**

# OBJETIVO DEL MODELO

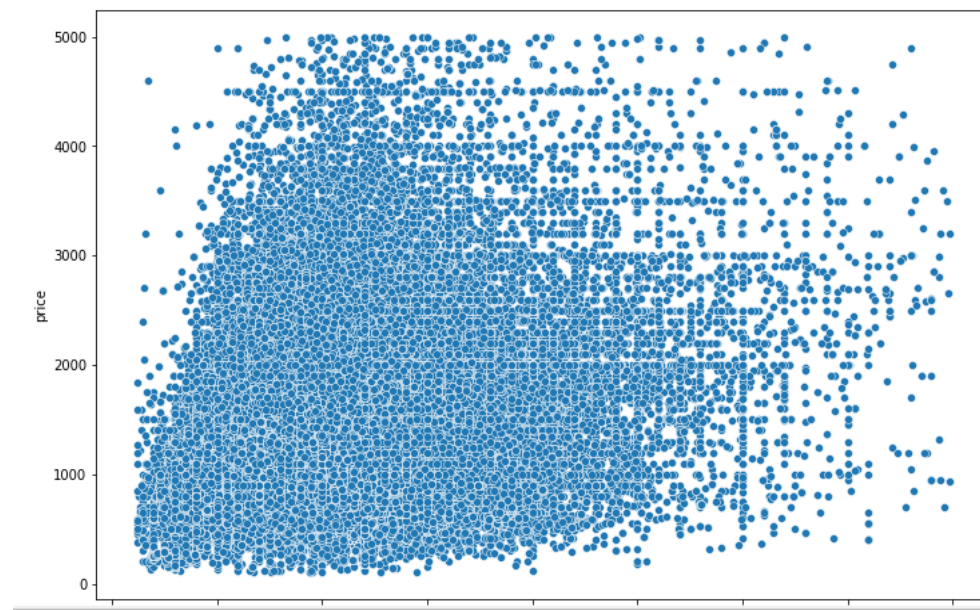
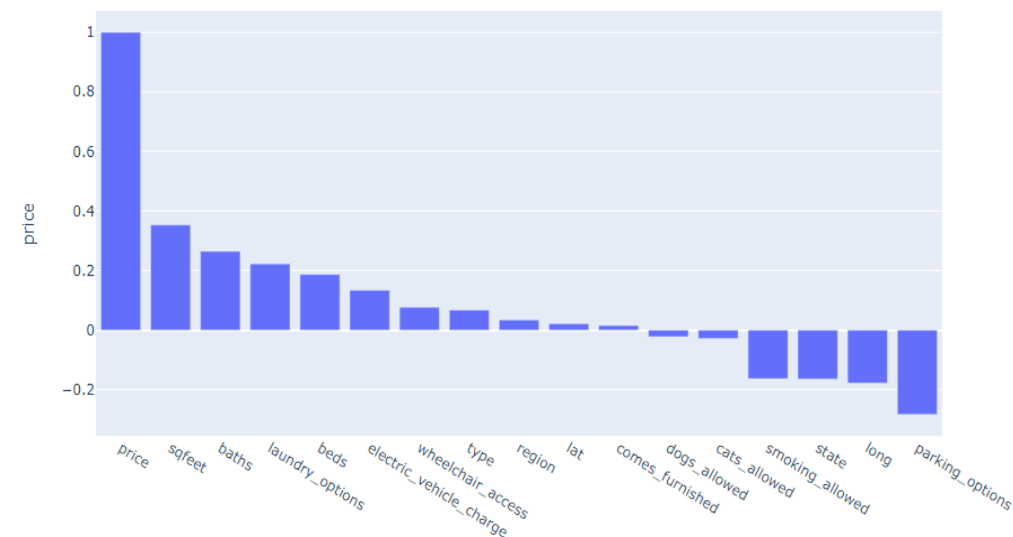
- PREDECIR EL LUGAR MEJOR UBICADO Y ACCESIBLE PARA QUE MI ESTADIA SEA LA MEJOR



# INSIGHTS

PODEMOS OBSERVAR QUE MIENTRAS MAS ALTO SEA EL PRECIO DEL LUGAR DONDE NOS HOSPEDEMOS , MEJORES SERVICIOS SE OFRECERAN. Y MIENTRAS MAS BAJOS SEAN LOS PRECIOS , SE COMPENSARA CON OTROS ITEMS COMO ESTACIONAMIENTO GRATIS

OTRO VALOR ENCONTRADO DONDE PODEMOS OBSERVAR QUE LA MAYORIA DE LOS HUESPEDES ELIGE DEPTOS O HABITACIONES CON UN PROMEDIO DE ALTURA EN MITAD DE EDIFICIO DE ACUERDO A SU PRECIO , LOS LUGARES MAS ALTOS Y CON MEJORES VISTAS SE ENCUENTRAN ENTRE LOS MAS COSTOSOS



# EL ALGORITMO QUE DECIDIMOS ELEGIR FUE DE REGRESION LINEAR

- DECIDIMOS UTILIZAR REGRESION LINEAR DEBIDO A QUE AL MOMENTO DE OPTIMIZARLO AGREGANDO MEJORAS , NOTAMOS QUE LOS DATOS FINALES ERAN
- MEJORES QUE LOS DATOS REALES QUE TENIAMOS
- La regresión lineal es una técnica de análisis de datos que predice el valor de datos desconocidos mediante el uso de otro valor de datos relacionado y conocido. Modela matemáticamente la variable desconocida o dependiente y la variable conocida o independiente como una ecuación lineal..

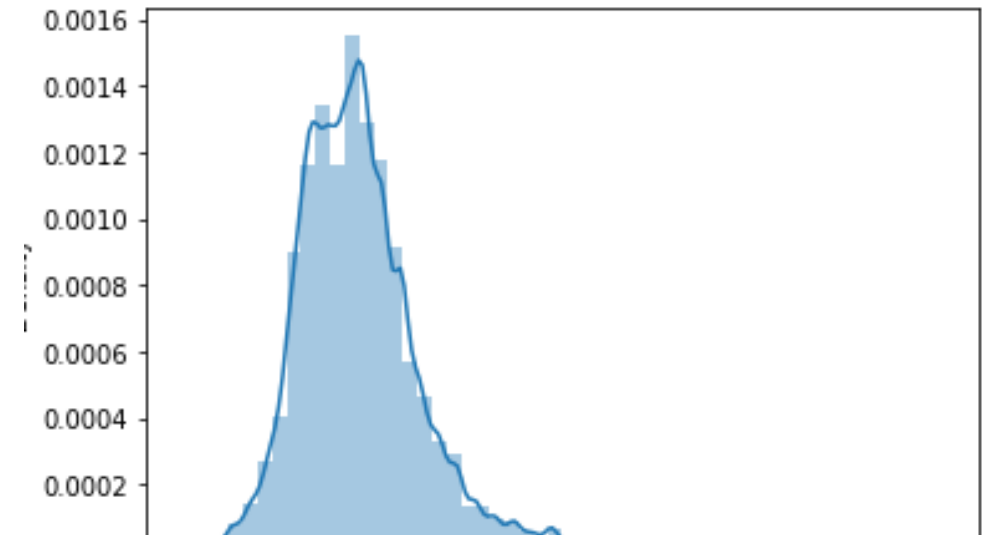
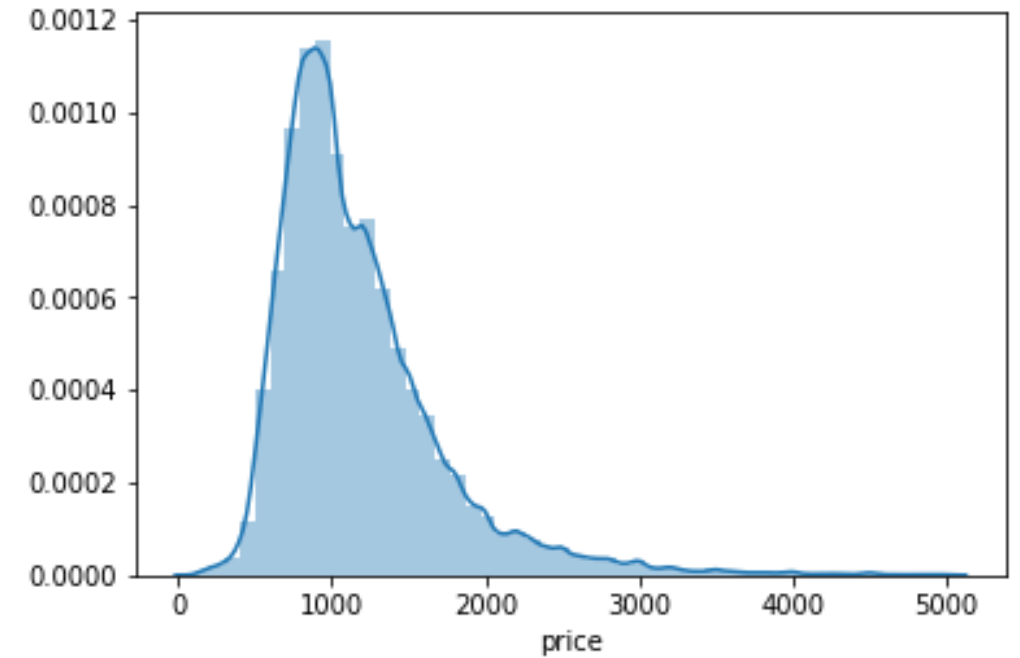
	True Values	Predicted Values
281334	880	995.975887
319393	1359	1495.187557
346212	1100	623.771214
44530	1575	1266.399124
157372	658	968.917620
...	...	...
330293	585	993.910633
175496	1550	1451.584744
361638	765	1086.332945
139465	3600	1156.943312
149905	1210	1249.483309



# Metricas

PODEMOS OBSERVAR QUE ANTES DE LAS MEJORAS,  
EL MODELO PRESENTABA UNA ASIMETRIA NEGATIVA  
, QUE ESTABA POR DEBAJO DEL 0.60

EN RELACION A PRECIO Y VISTAS EL MODELO NO ERA  
BUENO CON SUS METRICAS ORIGINALES , POR LO QUE SE  
PROCEDIO A REALIZAR MEJORAS Y A PROBAR MODELOS





# OPTIMIZACION

- OPTIMIZACION DEL MODELO : se fue mejorando el modelo mediante la utilizacion de variables ,
- originalmente el 2 score nos daba un 0.26 y al momento de mejorarlo con estas tecnicas obtuvimos un 0.80.
- Utilizamos 3 modelos para mejorar la prediccion y saber que modelo ayudaba mejor



	True Values	Predicted Values
281334	880	995.975887
319393	1359	1495.187557
346212	1100	623.771214
44530	1575	1266.399124
157372	658	968.917620
...	...	...
330293	585	993.910633
175496	1550	1451.584744
361638	765	1086.332945
139465	3600	1156.943312
149905	1210	1249.483309