

Traducción de variables (AirBnb)



**Tecnológico
de Monterrey**

A 16 de Febrero de 2025

Actividad 1 (Entendimiento de los Datos)

Alumno:

Braulio González Esquivel // A01425274

Analítica de datos y herramientas de inteligencia artificial I (Gpo 101)

Profesores: Juan Manuel Ahuactzin Larios,
Candy Yuridiana Alemán Muñoz,
Rigoberto Cerino Jiménez,
Alfredo García Suárez

CC: Sexto semestre profesional – Escuela de
Negocios – Licenciatura en Administración y
Transformación de Negocios.



Actividad 1

(Entendimiento de los Datos)



Índice

Contenido

Índice2

Introducción3

Elección del país y de la región4

Selección de las variables4

Traducción de las variables seleccionadas.....4

Conclusiones8

Bibliografía.....9

Introducción

La programación se ha convertido en una herramienta importante para las empresas que buscan extraer y manipular datos de manera constante. Al automatizar tareas, acceder a diversas fuentes de datos, transformar la información y realizar análisis avanzados, las empresas pueden obtener información valiosa que les permite avanzar y posicionarse de mejor manera frente a otras empresas.

Durante la extracción de información para su manipulación, es necesario depurar los aspectos de interés. Esto se logra con programación desde la base de los documentos.

En el siguiente trabajo académico se explica el proceso para llevar a cabo dicha extracción de datos.

Elección del país y de la región

Airbnb se ha posicionado como una de las empresas con mayor información de interés (bigdata). La empresa resurgió con fuerza. Redujo los números rojos a 352 millones en 2021 y en 2022 logró los primeros beneficios anuales de su historia, de 1.892 millones, tras crecer un 40% en ingresos, hasta 8.399 millones de dólares, más del doble que en 2020, el año de la pandemia. (Secretaría de Economía 2023).

A través de la información de consulta libre que la empresa nos proporciona a través de su página web, se ha realizado la elección de "New Brunswick, New Brunswick, Canada" debido a que resulta interesante la expansión de dicha provincia en el país de Canadá.

Selección de las variables

Dentro de la lista compuesta por 75 variables disponibles, se han seleccionado las que contienen datos enteros y de carácter booleano como prioridad, dejando en tercer lugar a las variables compuestas por datos condicionales, es decir, de varias opciones que varían dependiendo del dato introducido por el usuario.

Traducción de las variables seleccionadas

1: Identificador_ Identificador único de Airbnb para el anuncio

5: fuente_ Una de "búsqueda vecinal" o "raspado previo". "Búsqueda en el barrio" significa que el anuncio se encontró buscando en la ciudad, mientras que "búsqueda anterior" significa que el anuncio se vio en otra búsqueda realizada en los últimos 65 días y se confirmó que el anuncio seguía disponible en el sitio de Airbnb.

9: url de la imagen_ URL de la imagen de tamaño normal alojada en Airbnb para el anuncio

11: url del anfitrión_ La página de Airbnb del anfitrión

12: nombre del anfitrión_ Nombre del anfitrión. Normalmente sólo el nombre o nombres de pila.

13: anfitrión desde_ Fecha de creación del host/usuario. Para los anfitriones que son huéspedes de Airbnb, podría ser la fecha en la que se registraron como huéspedes.

20: url miniatura del anfitrión

21: url de la imagen del anfitrión

22: barrio del anfitrión

24: recuento total de listados del anfitrión_ El número de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb)

25: verificaciones del anfitrión

26: anfitrión tiene perfil imagen

27: Identidad verificada del anfitrión

28: Barrio

29: Depurado_ El barrio como geocodificado utilizando la latitud y la longitud frente a los barrios definidos por shapefiles digitales abiertos o públicos.

30: Depurado grupo del barrio_ El grupo de barrios se ha geocodificado utilizando la latitud y la longitud con respecto a los barrios definidos por archivos shape digitales abiertos o públicos.

31: Latitud_ Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para la latitud y la longitud.

32: Longitud_ Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para la latitud y la longitud.

33: Tipo de propiedad_ Tipo de propiedad seleccionada por el usuario. Los hoteles y bed and breakfast son descritos como tales por sus anfitriones en este campo.

34: Tipo de habitación_ [Casa entera/apartamento|Habitación privada|Habitación compartida|Hotel]

Todas las viviendas se agrupan en los tres tipos de habitación siguientes:

Casa entera

Habitación privada

Habitación compartida

Casa entera

Las habitaciones completas son las mejores si busca un hogar lejos de casa. Con un piso entero, tendrás todo el espacio para ti. Suele incluir un dormitorio, un cuarto de baño, una cocina y una entrada independiente. Los anfitriones deben indicar en la descripción si estarán en la propiedad o no (por ejemplo: "El anfitrión ocupa el primer piso de la casa"), y proporcionar más detalles en el anuncio.

Habitaciones privadas

Las habitaciones privadas son ideales cuando se prefiere un poco de intimidad, pero se valora la conexión local.

Cuando reserves una habitación privada, tendrás tu propia habitación privada para dormir y puede que compartas algunos espacios con otras personas. Es posible que tenga que caminar a través de espacios interiores que otro anfitrión o huésped puede ocupar para llegar a su habitación.

Habitaciones compartidas.

Las habitaciones compartidas son para cuando no te importa compartir un espacio con otras personas. Cuando reserves una habitación compartida, dormirás en un espacio compartido con otros y compartirás todo el espacio con otras personas. Las habitaciones compartidas son populares entre los viajeros flexibles que buscan nuevos amigos y estancias económicas.

35: Acomoda_ La capacidad máxima del listado

36: Baños_ El número de cuartos de baño del anuncio

37: Texto baños_ El número de cuartos de baño del anuncio.

En el sitio web de Airbnb, el campo de baños ha pasado de ser un número a ser una descripción textual. En los anuncios más antiguos se utiliza baños.

38: Habitaciones_ El número de habitaciones

39: Camas_ El número de cama(s)

40: Servicios o amenidades

41: Precio_ precio diario en moneda local.

NOTA: el signo \$ es un artefacto técnico de la exportación, por favor ignórelo.

42: Mínimo noches_ número mínimo de noches de estancia para el anuncio (las normas del calendario pueden ser diferentes)

43: Máximo noches_ número máximo de noches de estancia para el anuncio (las normas del calendario pueden ser diferentes)

44: Mínimo-mínimo noches_ el valor de noche_mínima más pequeño del calendario (365 noches en el futuro)

45: Máximo-mínimo noches_ el mayor valor de noche_mínima del calendario (365 noches en el futuro)

46: Mínimo-máximo noches_ el valor de noche_máxima más pequeño del calendario (365 noches en el futuro)

47: Máximo-máximo noches_ el mayor valor de noche_máxima del calendario (365 noches en el futuro)

48: Promedio noches mínimo_ el valor medio de noche_mínima del calendario (365 noches en el futuro)

49: Promedio noches máximo_ el valor medio de noche_máxima del calendario (365 noches en el futuro)

50: Calendario actualizado

51: Tiene disponibilidad_ [t=verdadero; f=falso]

- 52: Disponibilidad 30_ disponibilidad_x. La disponibilidad del anuncio x días en el futuro determinada por el calendario. Tenga en cuenta que un anuncio puede no estar disponible porque ha sido reservado por un huésped o bloqueado por el anfitrión.
- 53: Disponibilidad 60_ disponibilidad_x. La disponibilidad del anuncio x días en el futuro determinada por el calendario. Tenga en cuenta que un anuncio puede no estar disponible porque ha sido reservado por un huésped o bloqueado por el anfitrión.
- 54: Disponibilidad 90_ disponibilidad_x. La disponibilidad del anuncio x días en el futuro determinada por el calendario. Tenga en cuenta que un anuncio puede no estar disponible porque ha sido reservado por un huésped o bloqueado por el anfitrión.
- 55: Disponibilidad 365_ disponibilidad_x. La disponibilidad del anuncio x días en el futuro determinada por el calendario. Tenga en cuenta que un anuncio puede no estar disponible porque ha sido reservado por un huésped o bloqueado por el anfitrión.
- 56: último calendario alcanzado
- 57: Número de reseñas _ El número de reseñas que tiene el anuncio
- 58: Número de reseñas últimos 12 meses_ El número de reseñas que tiene el anuncio (en los últimos 12 meses)
- 59: Número de reseñas 30 días_ El número de reseñas que tiene el anuncio (en los últimos 30 días)
- 60: Primera reseña_ La fecha de la primera/antigua reseñas
- 71: Cálculo del recuento de las listas de anfitrión_ El número de listados que el anfitrión tiene en el scrape actual, en la geografía de la ciudad/región.
- 72: Cálculo del recuento de las listas de casas_ Número de anuncios de casas/apartamentos enteros que el anfitrión tiene en el scrape actual, en la ciudad/región geográfica
- 73: Cálculo del recuento de las listas de habitaciones privadas_ Número de anuncios de habitaciones privadas que el anfitrión tiene en el scrape actual, en la ciudad/región geográfica
- 74: Cálculo del recuento de las listas de habitaciones compartidas por el anfitrión_ Número de anuncios de habitaciones compartidas que el anfitrión tiene en el scrape actual, en la ciudad/región geográfica

Conclusiones

Bibliografía

AirBnb & Asociados, (2022), “Get the data” - “Inside Airbnb”
<https://insideairbnb.com/get-the-data/>

Miguel Jimenez,, (febrero, 2024) , “Airbnb planea reinventarse con la entrada en nuevos negocios tras lograr beneficios récord”
<https://elpais.com/economia/2024-02-13/airbnb-planea-reinventarse-con-la-entrada-en-nuevos-negocios-tras-lograr-beneficios-record.html#:~:text=Pero%20desde%20ese%20a%C3%B1o%20fat%C3%ADdico,el%20a%C3%B1o%20de%20la%20pandemia.>