

Alejandro vaqueiro vaqueiro

ester sánchez sánchez

inês bolaños pais

sílvia lendínez fernández

AIRBNB

Trabajo de fin de máster de big data & analytics

**Índice**

[Introducción 1](#_Toc4102233)

[1. Contexto 1](#_Toc4102234)

[2. Modelo de datos 5](#_Toc4102235)

# **Introducción**

Con la aparición de la economía colaborativa, surgen nuevos modelos de negocio que se basan en el uso de medios digitales para permitir la interacción entre un prestador de servicios y un usuario final. El concepto se basa en un consumo sostenible a través del reaprovechamiento de recursos, de la descentralización del poder y de un aumento de la calidad del servicio debido a una mayor oferta. Uno de los principales ejemplos de la economía colaborativa es Airbnb, la plataforma de alquiler de propiedades que pone en contacto anfitriones y huéspedes que, después de la experiencia, se valoran mutuamente para crear un sistema de referencia para futuros usuarios.

Aunque la empresa haya surgido en el contexto de la economía colaborativa, su rápido crecimiento y expansión a diversas ciudades del mundo ha hecho cuestionar los principios en los que se basa. Lo que inicialmente empezó como el reaprovechamiento de propiedades o habitaciones no usadas por los anfitriones, ha llevado a una explotación de propiedades de forma masiva, siendo necesaria la creación de una legislación específica o incluso la prohibición en muchas de las ciudades donde se instaló, para controlar el mercado inmobiliario. Este efecto ha generado movimientos en contra de la plataforma que cuestionan su verdadero impacto en las comunidades locales.

Uno de estos movimientos es *Inside Airbnb*, un proyecto que consiste en el web scrapping de los datos de Airbnb para que se puedan analizar de forma pública y evaluar el real impacto en las ciudades donde está presente. Con base en estos datos, este trabajo de fin de máster tiene como objetivo hacer un análisis descriptivo del impacto de la plataforma en las dos capitales de la Península Ibérica, Madrid y Lisboa, entender como está afectando el turismo y las comunidades locales y el cumplimiento con la legislación aplicada.

Para extracción y transformación de datos se usará la herramienta Python y para almacenamiento y análisis, SQL.

# **Contexto**

Con el objetivo de entender mejor el fenómeno del Airbnb, se contextualizará su surgimiento como forma de economía y consumo colaborativos, su desarrollo como empresa y los movimientos que han surgido en contra de su rápida expansión. Para el análisis posterior de las ciudades de Madrid y Lisboa, se explicará su evolución turística en los últimos años y el impacto del Airbnb en cada una de las ciudades.

**Economía colaborativa y consumo colaborativo**

El fenómeno de la economía colaborativa *(sharing economy)* ha crecido en los últimos años y deriva de *“desarrollos tecnológicos que han simplificado la posibilidad de compartir bienes y servicios tangibles e intangibles a través de la disponibilidad de varios sistemas de información en Internet”* (Hamari, Sjöklint, & Ukkonen, 2015). Este concepto está estrechamente relacionado con el consumo colaborativo *(collaborative consumption)*, que consiste en obtener o dar acceso a bienes y servicios a través de comunidades online (Hamari, Sjöklint, & Ukkonen, 2015). En el caso de los servicios, la etapa posterior a la transacción es importante para captación de valor por parte de los proveedores. De esta forma, variables como *reviews* bidireccionales o respuestas a los usuarios ganan cada vez más importancia en páginas web de economía colaborativa (Zervas, Proserpio, & Byers, 2017).

Los principales ejemplos de economía colaborativa son las plataformas Airbnb, para alquiler de alojamiento turístico y Uber, una plataforma de movilidad. En ambos casos, han empezado como startups de Silicon Valley que han tenido un rápido crecimiento y actualmente son multinacionales que operan en ciudades por todo el mundo y se estima su valoración en miles de millones de dólares (Salter, 2012; Hartmans, 2017).

**Airbnb: de startup a multinacional**

En el caso de Airbnb, la empresa surgió en 2008 cuando dos estudiantes de San Francisco encontraron ocupados todos los hoteles de su ciudad con motivo de una feria. De esa forma, decidieron crear un sitio web que alquilaría las camas hinchables que les sobraban en el apartamento para que más gente pudiese tener un alojamiento. Tras una pequeña financiación de 20.000$ por Y-Combinator, una aceleradora de startups, el fondo de inversión de capital riesgo Sequoia Capital, impulsora de empresas como Apple, Google, Yahoo y Oracle, la financió con 600.000$ (Álvarez, 2016).

La empresa ha crecido de una forma exponencial en los últimos años, especialmente en las grandes capitales europeas. Pese a ello, sigue teniendo capital privado y no ha salido a bolsa a fecha de marzo de 2019. A finales del año 2015 ya estaba presente en 191 países y 34.000 ciudades. El impacto originado en las grandes ciudades ha desencadenado procesos regulatorios para controlar su uso y asegurar que las viviendas se alquilan en los términos de alojamiento turístico (Miltra, 2016).

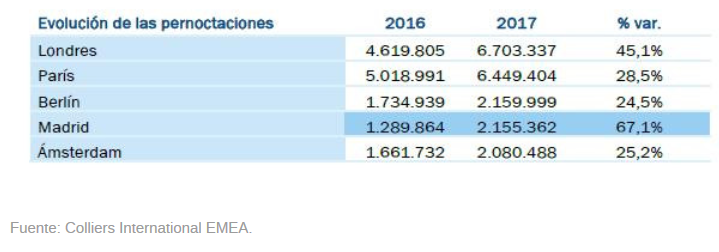
**Los movimientos anti Airbnb** (work in progress)

**La evolución del turismo y del Airbnb en Madrid y Lisboa** (work in progress)

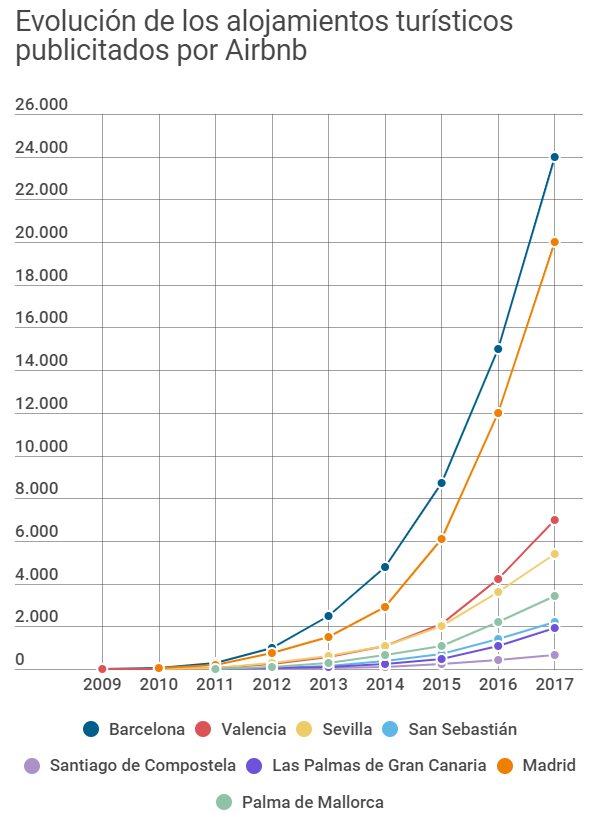
Aunque tanto Madrid como Lisboa sean las dos capitales de la Península Ibérica, las ciudades son muy distintas a nivel de turismo.

**Madrid**

La cuota de mercado en pernoctaciones ha crecido desde 2016, llegando a los dos millones de estancias. Como vemos en la tabla 1, Madrid es la ciudad europea que más ha crecido, frente al resto de capitales europeas, alrededor de un 67%.



A nivel de alojamientos turísticos, ha tenido un crecimiento desde 2009, sin embargo, quedándose por detrás de Barcelona. En 2017 ha alcanzado los 20.000 alojamientos anunciados en la plataforma, 8.000 nuevas viviendas más que en el año anterior.



<https://www.hosteltur.com/127379_8-ciudades-espana-donde-ha-crecido-airbnb.html>

La cuota de mercado en pernoctaciones ha crecido desde 2016 en Madrid, llegando a los dos millones de estancias. Como vemos en esta tabla, Madrid es la ciudad europea que más ha crecido, frente al resto de capitales europeas, alrededor de un 67%.

Por otra parte, la actividad de Airbnb en Madrid está muy concentrada en el Distrito Centro, que supone un 61% de los alojamientos reservados. Mientras que, en los distritos de las ciudades europeas con más peso, la media está entre el 25 y 30%. De hecho, los propietarios de estas viviendas tienen una rentabilidad mayor que los establecimientos hosteleros de estas ciudades

http://www.aept.org/archivos/documentos/ostelea\_informe\_economia\_colaborativa.pdf

Las tarifas medias diarias más baratas de Europa se encuentran entre otras capitales, en Madrid. Se pueden encontrar viviendas por menos de 50€ y los Airbnb de gama más alta también tienen precios más reducidos.

**Lisboa**

El turismo en Portugal viene aumentando desde 2012, superando en 2017 los 20 millones de huéspedes (gráfico 1) y 57 millones de dormidas (gráfico Y en anexo) y con un crecimiento medio de 7% para el período.

Gráfico 1 – Evolución del total de huéspedes al año en Portugal

Lisboa en específico, ha sido en 2017 la segunda ciudad europea con mayor crecimiento turístico (10,2%), considerando el período 2009 – 2016, según el Global Destination Index.

<https://newsroom.mastercard.com/wp-content/uploads/2017/10/Mastercard-Destination-Cities-Index-Deck.pdf>

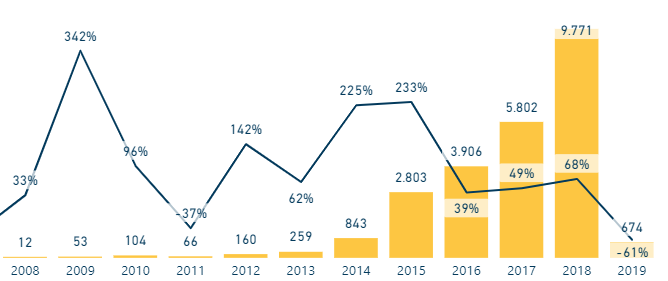
Un conjunto de políticas implementadas entre los años 2008 y 2012, para recuperar de la crisis financiera y de la intervención del Fondo Monetario Internacional en el país, han impactado en la liberalización del mercado inmobiliario en Portugal. Medidas como beneficios fiscales, facilidades administrativas y visados a inversores externos, junto con proyectos de rehabilitación de la ciudad, han convertido al país en un atractivo financiero inmobiliario (Calvo & Ramos, 2018).

Con el crecimiento del turismo en los últimos años, ha aparecido en 2008 el concepto legal de “alojamiento local” (decreto ley 39/2008), que se definía como las propiedades que ofrecían servicios de alquiler temporal a través de una remuneración, pero sin los requisitos para que se consideraran como empresas turísticas. Desde este momento que hay una obligación de registrar las propiedades en el ayuntamiento local, garantizar los requisitos mínimos de higiene y seguridad y que el servicio de alquiler no supere los 30 días. El límite de huéspedes está definido por las camas disponibles en la propiedad.

En 2014, debido a la importancia turística que ha ganado este tipo de propiedades, se ha creado una legislación específica (decreto ley 128/2014), que impone un límite de explotación a los propietarios de máximo 9 apartamentos por edificio, excepto si estos representan más del 75% del número de apartamentos del edificio (excepción implementada en el decreto ley 63/2015). Se impone aún la capacidad máxima de 9 habitaciones y 30 huéspedes por propiedad. Se estrechan las condiciones de seguridad, obligando a la propiedad que disponga de un extintor, manta de incendios y un maletín de primeros auxilios. A nivel del registro, este pasa a ser obligatorio de forma online en el Turismo de Portugal.

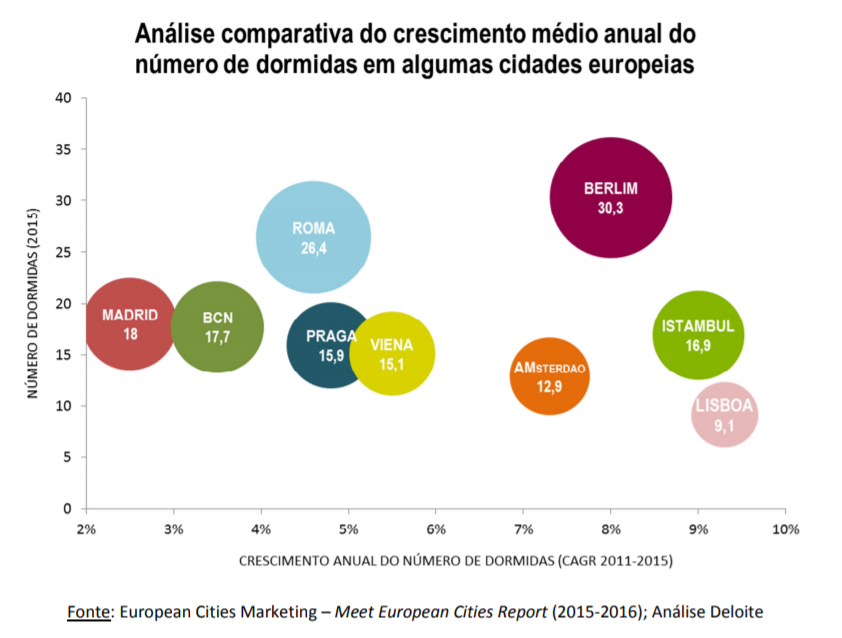
En 2018, sale una nueva ley (decreto ley 62/2018) que limita el alquiler de habitaciones a la residencia del propietario, que tiene que corresponder a su domicilio fiscal y donde puede alquilar como máximo 3 habitaciones. Se introducen también cambios al límite de camas, que no pueden superar el doble de habitaciones, más dos camas extra y dos camas suplementares para niños hasta 12 años.

En noviembre de 2018, debido a la concentración de alojamientos locales en el centro histórico de la ciudad, el ayuntamiento de Lisboa suspende durante un año nuevos registros de estas propiedades en las zonas turísticas de la ciudad. En concreto, en los barrios: Bairro Alto, Madragoa, Castelo, Alfama y Mouraria (Boletim Câmara Municipal de Lisboa, nº 1293).



<http://www.cm-lisboa.pt/fileadmin/VIVER/Urbanismo/urbanismo/BM_28nov2018_1Supl.pdf>

<http://www.cm-lisboa.pt/fileadmin/VIVER/Urbanismo/urbanismo/estudos/turismo/Estudo__ZTH_Lx_16102018.pdf>



# **Modelo de datos**

El modelo de datos consiste en la extracción, transformación y almacenamiento de los datos en una base de datos.

* 1. **Extracción**

Airbnb no comparte sus datos de forma pública. Tiene una API de acceso limitado para ayudar a los anfitriones a gestionar sus reservas y anuncios y crear flujos de trabajo, pero no permite acceder a información sobre propiedades disponibles por región, propietarios o alquileres. Sin embargo, dentro del movimiento anti Airbnb, ha surgido el proyecto *Inside Airbnb,* creado por el activista australiano Murray Cox, para evaluar el impacto del uso de la plataforma en su país natal. Algunos ayuntamientos del país han incluso utilizado estos datos en vez de los oficiales de la plataforma para analizar el impacto local. El activista dice que el gobierno de Australia ha permitido que la empresa se implantara en la ciudad para pasar una imagen imagen liberal y adaptada a la nueva realidad, pero sin evaluar su impacto en las rentas y en la infraestructura. De esta forma, el proyecto ha sido creado con el objetivo de entender hasta qué punto se está aplicando el concepto de economía compartida, debido a un posible uso de alojamientos que serían de habitación primaria para alquiler exclusivamente turístico, impactando así la vivienda y las comunidades. Airbnb ha criticado al proyecto por información incorrecta y engañosa, pero numerosos estudios, incluidos gobiernos locales, se han basado en estos datos para estimar el impacto de Airbnb en las ciudades (Han, 2018).

Los datos de *Inside Airbnb* se han obtenido a través de web scrapping de la página web pública de Airbnb para un determinado momento en el tiempo. Estos datos han sido verificados, limpios, analizados y agregados por el equipo de *Inside Airbnb.* No se comparte información privada. Toda la información está disponible de forma pública en la página de Airbnb (Inside Airbnb, 2019).

De esta forma, los datos que sirven de base al trabajo se han extraído de esta página para las ciudades de Madrid y Lisboa para el período de XXX. Para cada ciudad y período, se ha extraído el archivo *listings.csv.gz* que contiene los datos sin agrupar, correspondientes a una fecha en específico, con periodicidad mensual. Se han considerado también los archivos *neighbourhoods.geojson* y *neighbourhoods.csv* que permiten geolocalizar los apartamentos en el mapa.

Estos archivos se han importado en un notebook de Jupyter y han sido transformados en Python.

* 1. **Transformación**

Una vez importados los archivos, se ha confirmado la uniformización en el nombre de columnas, todas en minúsculas y sin espacios.

En una primera etapa, se definen las variables no relevantes para el análisis del total de las 106 iniciales. Las variables se han eliminado según los siguientes criterios:

* Variables con campos de texto sin relevancia, no predeterminados por Airbnb y, por lo tanto, sin un formato establecido. Según este criterio, se han eliminado 6 variables: *host\_name, summary, space, neighbourhood\_overview, notes, transit, access, interaction, house\_rules*
* Variables redundantes, cuya información se puede obtener con otras variables. Según este criterio, se han eliminado 11 variables: *scrape\_id, calendar\_last\_scraped, jurisdiction\_names, country, country\_code, street, market, smart\_location, state, zipcode, city*
* Variables que corresponden a url’s no relevantes para el análisis. Según este criterio, se han eliminado 6 variables: *thumbnail\_url, picture\_url, medium\_url, xl\_picture\_url, host\_thumbnail\_url, host\_picture\_url*
* Variables cuyo método de cálculo se desconoce o cuyo significado no se comprende. Según este criterio, se han eliminado 9 variables: *is\_location\_exact, maximum\_maximum\_nights, maximum\_minimum\_nights, mínimum\_minimum\_nights, minimum\_maximum\_nights, is\_business\_travel\_ready, hsot\_verifications, host\_identity\_verified*
* Variables que hacen referencia al método de scrapping o que están condicionadas por el momento concreto en que se hizo. Según este criterio, se han eliminado 4 variables: *maximum\_nights, minimum\_nights, calendar\_updated, has\_availability*
* Variables cuya información no es necesaria para el análisis. Según este criterio, se han eliminado 2 variables: *require\_guest\_profile\_picture, require\_guest\_phone\_verification*.

Después de este proceso, sigue habiendo un total de 68 variables. Para un análisis fiable y comparable entre ciudades y períodos de tiempo, eliminamos las variables con un porcentaje de nulos superior al 30%, ya que no serán tan útiles con tanta información perdida. De esta forma, se han eliminado YY variables más:

Teniendo ya el número final de variables, se convierten las que quedan para un análisis más eficiente:

* Se cambia de formato texto a formato fecha las siguientes variables: *last\_scraped, host\_since, first\_review, last\_review*
* Se reemplazan valores de *true* y *false* por 1 y 0 en las siguientes variables: *host\_is\_superhost, requires\_license, instant\_bookable*
* Se cambia a categórica la variable *cancellation\_policy* según las políticas de cancelación de Airbnb (anexo): *flexible, moderate, strict\_14\_with\_grace\_period, super\_strict\_30, super\_strict\_60*
* Se crea una nueva columna con la ciudad a la que se refieren los datos (Lisboa, Madrid)

De esta forma, la base de datos final tiene un total de XX variables.

* 1. **Almacenamiento**

Una vez teniendo el total de variables, se definen las tablas a crear, agrupando las variables por categorías.

|  |  |  |
| --- | --- | --- |
| Variable | tipo | descripción |
| *id* |  | Id de la propiedad |
| *listing\_url* |  | url de la propiedad |
| last\_scrapped |  | Fecha de scrapping de los datos |
| name |  | Nombre del anfitrión |
| description |  | Descripción de la propiedad |
| host\_id |  | Id del anfitrión |
| host\_since |  | Fecha de inicio como anfitrión |
| host\_is\_superhost |  | Anfitrión es o no superhost (anexo) |
| host\_neighbourhood |  | Localización del anfitrión |
| host\_listings\_count |  | Total de propiedades que tiene el huésped en Airbnb |
| host\_total\_listings\_count |  | Total de propiedades que tiene el huésped en Airbnb |
| neighbourhood\_cleansed |  |  |
| neighbourhood\_group\_cleansed |  |  |
| city |  | Ciudad en análisis |
| latitude |  | Latitud de la propiedad (de 0 a 150m de la localización real) |
| longitude |  | Longitud de la propiedad (de 0 a 150m de la localización real) |
| property\_type |  | Tipo de propiedad |
| room\_type |  | Tipo de habitación |
| accommodates |  | Número máximo de huéspedes |
| bathrooms |  | Número de baños (puede ser 0,5 si no incluye ducha) |
| bedrooms |  | Número de habitaciones |
| beds |  | Número de camas |
| bed\_type |  | Tipo de cama |
| amenities |  | Comodidades de la propiedad (predefinidas por Airbnb) |
| price |  | Precio |
| cleaning\_fee |  | Tasa de limpieza |
| guests\_included |  | Número máximo de huéspedes para el precio fijo |
| extra\_people |  | Precio de cada persona extra a guests\_included |
| minimum\_nights\_avg\_ntm |  | ¿? |
| maximum\_nights\_avg\_ntm |  | ¿? |
| availability\_30 |  |  |
| availability\_60 |  |  |
| availability\_90 |  |  |
| availability\_365 |  |  |
| number\_of\_reviews |  | Número total de reviews |
| number\_of\_reviews\_ltm |  | Número de reviews en los últimos 12 meses |
| first\_review |  | Fecha de la primera review |
| last\_review |  | Fecha de la última review |
| review\_scores\_rating |  |  |
| review\_scores\_accuracy |  |  |
| review\_scores\_cleanliness |  |  |
| review\_scores\_checkin |  |  |
| review\_scores\_communication |  |  |
| review\_scores\_location |  |  |
| review\_scores\_value |  |  |
| requires\_license |  |  |
| instant\_bookable |  |  |
| cancellation\_policy |  |  |
| 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 |  |  |

Explicar las variables: precio fijo y variable, reviews y Sistema de reviews, superhost, listings\_count,.

Una limitación de los datos extraídos de *Inside Airbnb* tiene que ver con la dificultad en estimar la tasa de ocupación de las propiedades listadas y los ingresos del anfitrión, para poder calcular la rentabilidad del alquiler. Sin embargo, la propia página sugiere un método de cálculo que usa las reviews como proxy de reservas estimadas. Este se denomina “Modelo de San Francisco”, por basarse en dos modelos creados para la ciudad para cuantificar el impacto de Airbnb a nivel de políticas públicas y de planeamiento urbano en las viviendas. El primer modelo, creado por Alex Marqusee, estima que las reviews representan el 72% de los alquileres, aunque no se considera una fuente fiable. El segundo modelo, del *Budget and Legislative Analyst’s Office,* indica una representatividad del 30,5%, basándose en la comparación de datos públicos con los datos del reporte de Airbnb de 2014. Sin embargo, la página recomienda el uso de un modelo intermedio, que asume una tasa del 50%.

Usando los datos de la media de la estancia y multiplicándolos por los alquileres estimados, se puede obtener la tasa de ocupación.

* An **average length of stay** is configured for each city, and this, multiplied by the**estimated bookings** for each listing over a period gives the **occupancy rate**
  + Where statements have been made about the average length of stay of Airbnb guests for a city, this was used.
  + For example, [Airbnb reported 5.5 nights](http://blog.airbnb.com/economic-impact-airbnb/#san-francisco) as the average length of stay for guests using Airbnb in San Francisco.
  + Where no public statements were made about average stays, a value of **3 nights per booking** was used.
  + If a listing has a **higher minimum nights** value than the average length of stay, the minimum nights value was used instead.
* The **occupancy rate** was **capped at 70%** - a relatively high, but reasonable number for a highly occupied "hotel".
  + This controls for situations where an Airbnb host might change their minimum nights during the high season, without the review data having a chance to catch up; or for a listing with a very high review rate.
  + It also ensures that the occupancy model remains conservative.
* **Number of nights** booked or available per year for the **high availability** and **frequently rented** metrics and filters were generally aligned with a city's short term rental laws designed to **protect residential housing**.
* Location information for listings are anonymized by Airbnb.
  + In practice, this means the location for a listing on the map, or in the data will be from 0-450 feet (150 metres) of the actual address.
  + Listings in the same building are anonymized by Airbnb individually, and therefore may appear "scattered" in the area surrounding the actual address.
* Listings can be deleted in the Airbnb platform. The data presented here is a snapshot of listings available at a particular time. Other snapshots of data from previous dates are available for analysis by request, and a future activity may be to analyse the characteristics of deleted Airbnb listings.
* The Airbnb calendar for a listing does not differentiate between a booked night vs an unavailable night, therefore these bookings have been counted as "unavailable". This serves to understate the Availability metric because popular listings will be "booked" rather than being "blacked out" by a host.
* Some hosts might not keep their calendar updated, or have it highly available even though they live in the entire home/apartment.
  + To ensure you are only seeing listings that are both "highly available" and being booked frequently, use the "Only highly available" filter with the "Only recent and frequently booked" filter to only select listings that have only been rented recently (reviewed in the last 6 months), and are being rented regularly (number of nights per year greater than the threshold for that city).
* Some reviews may be "spam" allowed by Airbnb. Analysis suggests that spam reviews are small and do not affect the statistics.
* Neighbourhood names for each listing are compiled by comparing the listing's geographic coordinates with a city's definition of neighbourhoods. Airbnb neighbourhood names are not used because of their inaccuracies.
* Extracción (fuente de datos): airbnb get the data, web scrapping
* Transformación (tratamiento y validación de datos)
* Almacenamiento

Visualización

* Análisis descriptivo
* Análisis predictivo

**Bibliografía**

Hamari, J., Sjöklint, M., & Ukkonen, A. (2015). The Sharing Economy: Why People Participate in Collaborative Consumption. *Journal of the Association for Information Science and Technology, 67*(9), pp. 2047-2059.

Han, M. (26 de febrero de 2018). Murray Cox, the Australian 'data activist' taking on Airbnb. *Financial Review*. Obtido em 20 de marzo de 2019, de https://www.afr.com/real-estate/residential/murray-cox-the-australian-data-activist-taking-on-airbnb-20180226-h0wn4g

Hartmans, A. (4 de mayo de 2017). Uber CEO Travis Kalanick's Super Successful and Controversial Life Story. *Inc.* Obtido em 20 de marzo de 2019, de https://www.inc.com/business-insider/uber-ceo-travis-kalanick-life-story.html

Inside Airbnb. (2019). *About*. Obtido em 20 de marzo de 2019, de Inside Airbnb: http://insideairbnb.com/about.html

Salter, J. (07 de septiembre de 2012). Airbnb: The story behind the $1.3bn room-letting website. *The Telegraph*. Obtido em 20 de marzo de 2019, de https://www.telegraph.co.uk/technology/news/9525267/Airbnb-The-story-behind-the-1.3bn-room-letting-website.html

Zervas, G., Proserpio, D., & Byers, J. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research, 54*(5), pp. 687-705.