



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Politécnica Superior de Alcoy

Data analytics para la mejora en el proceso de toma de
decisiones

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Beltrán Dolz, Raúl

Tutor/a: Sanchis Gisbert, Raquel

CURSO ACADÉMICO: 2021/2022

RESUMEN

En el contexto empresarial actual, en el que impera una gran incertidumbre debido a los acontecimientos que últimamente han ocurrido, cobra especial importancia el proceso de toma de decisiones para poder discernir las alternativas más adecuadas en cada situación. Es por ello que uno de los enfoques actuales que pueden ayudar a mejorar el proceso de toma de decisiones es el *data analytics* (análisis de datos). Dicho enfoque, como su propio nombre indica, analiza datos con el objetivo de para tratar de encontrar tendencias y métricas para que las empresas puedan tomar decisiones más acertadas. Es por ello que el principal objetivo del presente Trabajo Fin de Grado (TFG) es la utilización del análisis de datos para tomar decisiones que permitan obtener una rápida visualización de todos los datos necesarios para tener una visión amplia de ellos. Con la visualización de datos histórica se identificarán patrones de una manera más sencilla y de ahí ver las tendencias que puede seguir un activo, de forma que se mejore y optimice el proceso de toma de decisiones en cualquier ámbito.

PALABRAS CLAVE

Data analytics; toma de decisiones; datos; tendencia; patrón.

ABSTRACT

In the current business context, in which there is a great deal of uncertainty due to recent events, the decision-making process is of particular importance in order to be able to discern the most appropriate alternatives in each situation. That is why one of the current approaches that can help to improve the decision-making process is data analytics. This approach, as its name suggests, analyses data in order to try to find trends and metrics so that companies can make better decisions. That is why the main objective of this Final Degree Project (TFG) is the use of data analytics to make decisions that allow a quick visualisation of all the data necessary to have a broad view of them. With the visualisation of historical data, patterns will be identified in a simpler way and from there, it will be possible to see the trends that an asset may follow, in order to improve and optimise the decision-making process in any field.

KEYWORDS

Data analytics; decision making; data; trend; pattern.



Índice

Índice.....	2
1. Introducción.....	7
2. Objetivos y metodología.....	11
3. Caso de Estudio.....	14
3.1. Airbnb.....	14
3.1.1 Historia.....	14
3.1.2 La empresa.....	14
3.1.3 Motivación de la elección de Airbnb.....	15
3.2. Obtención del conjunto de datos.....	16
3.3. Procesado y limpieza de datos.....	17
3.3.1. Problemas con la limpieza de los datos.....	17
3.3.2 Procesado de datos.....	18
3.4. Análisis de datos.....	19
3.4.1 Tipo de alquileres.....	19
3.4.2 Hosts.....	26
3.4.3 Género.....	35
3.4.4 Barrios.....	40
3.4.5 Reseñas.....	57
3.5. Detección y propuesta de mejoras.....	62
3.5.1 En los tipos de alquiler.....	62
3.5.2 En los hosts.....	63
3.5.3 Por género.....	64
3.5.4 Por barrios.....	65
3.5.5 En las reseñas.....	66
4. Conclusiones.....	67
Referencias.....	69

Índice de Figuras

Figura 1. Representación simplificada del ciclo del análisis de datos (Fuente: https://www.uipath.com)	9
Figura 2. Metodología del TFG (Fuente: Elaboración propia).....	12
Figura 3. Airbnb en el año 2010 (Fuente: www.webdesignmuseum.org).....	14
Figura 4. Habitación privada en una isla (Fuente: www.airbnb.es)	15
Figura 5. Representación por sub-barrios de cantidad de cada tipo de alquileres (Fuente: Elaboración propia)	21
Figura 6. Cantidad media de reseñas por tipo de alquiler (Fuente: Elaboración propia).....	23
Figura 7. Representación de la media de todas las reseñas por tipo de alquiler (Fuente: Elaboración propia)	24
Figura 8. Correlación entre las distintas reseñas (Fuente: Elaboración propia)	25
Figura 9. Página Airbnb apartamento con servicios (Fuente: Airbnb.es)	27
Figura 10. Cuenta de usuario Airbnb (Fuente: Elaboración propia)	27
Figura 11. Gráfica de días medios que lleva el host en Airbnb por el ratio de respuesta (Fuente: Elaboración propia)	30
Figura 12. Gráfica de días medios que lleva el host en Airbnb por el ratio de aceptación (Fuente: Elaboración propia)	32
Figura 13. Gráfica de días medios que lleva el host en Airbnb por su condición de superhost (Fuente: Elaboración propia).....	32
Figura 14. Gráfica de la relación entre géneros. (Fuente: Elaboración propia).....	38
Figura 15. Gráfica de la cantidad media de casas por género (Fuente: Elaboración propia).....	39
Figura 16. Barrios de Valencia (Fuente: www.valencia.es)	40
Figura 17. Correlación entre habitaciones, camas y el precio (Fuente: Elaboración propia).....	41
Figura 18. Gráfica de la cantidad de licencias que hay por barrio (Fuente: Elaboración propia).....	42
Figura 19. Correlación entre el total de licencias y el total de alquileres (Fuente: Elaboración propia)	44
Figura 20. Diagrama de caja y bigotes de las disponibilidades de 30 y 60 días (Fuente: Elaboración propia)	45
Figura 21. Diagrama de caja y bigotes de las disponibilidades de 90 y 365 días (Fuente: Elaboración propia)	45
Figura 22. Diagrama de caja y bigotes de la disponibilidad de 30 días por barrios (Fuente: Elaboración propia)	46
Figura 23. Diagrama de caja y bigotes de la disponibilidad de 60 días por barrios (Fuente: Elaboración propia)	47
Figura 24. Diagrama de caja y bigotes de la disponibilidad de 90 días por barrios (Fuente: Elaboración propia)	47

Figura 25. Diagrama de caja y bigotes de la disponibilidad de 365 días por barrios (Fuente: Elaboración propia)	48
Figura 26. Gráfica de la tendencia de disponibilidad en el ultimo año (Fuente: Elaboración propia)	50
Figura 27. Gráfico de la disponibilidad de 30 días por barrios (Fuente: Elaboración propia).....	50
Figura 28. Gráfico de la disponibilidad de 60 días por barrios (Fuente: Elaboración propia).....	50
Figura 29. Gráfico de la disponibilidad de 90 días por barrios (Fuente: Elaboración propia).....	51
Figura 30. Gráfico de la disponibilidad de 365 días por barrios (Fuente: Elaboración propia).....	51
Figura 31. Representación en mapa de las disponibilidades medias de 30 y 60 días por barrios (Fuente: Elaboración propia).....	52
Figura 32. Representación en mapa de las disponibilidades medias de 90 y 365 días por barrios (Fuente: Elaboración propia).....	52
Figura 33. Correlación entre camas y precios (Fuente: Elaboración propia)	52
Figura 34. Representación de la correlación entre el precio y las distintas disponibilidades (Fuente: Elaboración propia).....	53
Figura 35. Representación de la correlación entre las camas y las distintas disponibilidades (Fuente: Elaboración propia).....	54
Figura 36. Representación gráfica de las reseñas de cada idioma en el año 2018 (Fuente: Elaboración propia)	60
Figura 37. Representación gráfica de las reseñas de cada idioma en el año 2019 (Fuente: Elaboración propia)	61
Figura 38. Representación gráfica de las reseñas de cada idioma en el año 2020 (Fuente: Elaboración propia)	61
Figura 39. Representación gráfica de las reseñas de cada idioma en el año 2021 (Fuente: Elaboración propia)	61
Figura 40. Representación de la correlación entre el precio y el número de reseñas (Fuente: Elaboración propia)	61
Figura 41. Representación de la correlación entre el precio y la valoración media de reseñas (Fuente: Elaboración propia)	62



Índice de Tablas

Tabla 1: Cantidad total de alquileres por tipo. Fuente: Elaboración propia	19
Tabla 2: porcentaje total de alquileres por tipo. Fuente: Elaboración propia.....	20
Tabla 3: Cantidad de alquileres de cada tipo en cada barrio. Fuente: Elaboracion propia	20
Tabla 4:Cantidad de reseñas y sus valores por tipo de alquiler. Fuente: Elaboración Propia.....	22
Tabla 5: Valor reseña global y cantidad media de reseñas por tipo de apartamento. Fuente: Elaboración propia.....	22
Tabla 6: Valor media de todas las reseñas por tipo de alquiler. Fuente: Elaboración propia	23
Tabla 7: Valor media de todas las reseñas por tipo de alquiler. Fuente: Elaboración propia	26
Tabla 8:Ratio de respuesta y de aceptación en función del superhost y licencia. Fuente: Elaboración propia.....	28
Tabla 9 :Ratio de respuesta y de aceptación en función de superhost y licencia con fecha en días. Fuente: Elaboración propia	29
Tabla 10: Días medios del host en Airbnb agrupado por ratio de respuesta. Fuente: Elaboración propia.....	29
Tabla 11: Cantidad de host que responden por cada ratio de respuesta. Fuente: Elaboración propia.....	30
Tabla 12: Días medios del host en Airbnb agrupado por ratio de aceptación. Fuente: Elaboración propia.....	31
Tabla 13:Cantidad de alquileres, reseñas valor medio de reseñas por género y por tipo de superhost de casas enteras. Fuente: Elaboración propia	33
Tabla 14:Cantidad de alquileres, reseñas valor medio de reseñas por género y por tipo de superhost de habitaciones privadas. Fuente: Elaboración propia	33
Tabla 15: Cantidad de alquileres, reseñas valor medio de reseñas por género y por tipo de superhost de habitaciones compartidas. Fuente: Elaboración propia	35
Tabla 16: Género del host y licencia. Fuente: Elaboración propia.....	35
Tabla 17: Cantidad de host y licencias por genero en casas enteras . Fuente: Elaboración propia	35
Tabla 18: Cantidad de host y licencias por genero en habitaciones privadas. Fuente: Elaboración propia.....	36
Tabla 19: Cantidad de host y licencias por genero en habitaciones compartidas . Fuente: Elaboración propia.....	36
Tabla 20: Nombre del host junto con su género y la cantidad de alquileres que tiene de cada tipo. Fuente: Elaboración propia	37
Tabla 21: Cantidad de alquileres que tiene de casa complicada host agrupado por género. Fuente: Elaboración propia.....	37
Tabla 22: Cantidad de habitaciones, camas y precio medio agrupado por barrio. Fuente: Elaboración propia.....	41



Tabla 23: Cantidad de licencias agrupado por barrio. Fuente: Elaboración propia	42
Tabla 24: Cantidad de alquileres agrupado por barrio. Fuente: Elaboración propia.....	43
Tabla 25: Cantidad de alquileres, licencias y el porcentaje de estas agrupado por barrio. Fuente: Elaboración propia.....	43
Tabla 26: Cantidad de camas, su disponibilidad y el precio. Fuente: Elaboración propia	44
Tabla 27: Disponibilidades absolutas y porcentuales agrupadas por barrio. Fuente: Elaboración propia.....	48
Tabla 28: Cantidad de camas en alojamientos con 1 habitación. Elaboración propia....	54
Tabla 29: Cantidad de camas en alojamientos con 2 habitaciones. Fuente: Elaboración propia	54
Tabla 30: Reseñas, disponibilidades y zona del barrio agrupado por barrio. Fuente: Elaboración propia.....	55
Tabla 31: Cantidad media de reseñas, disponibilidades y precio medio por cama agrupado por zona. Fuente: Elaboración propia	56
Tabla 32: Ingresos medios agrupados por zona. Fuente: Elaboración propia	57
Tabla 33: Id del host, fecha, nombre del usuario y reseña. Fuente: Elaboración propia	57
Tabla 34: Traducción nombres de idioma a ISO-639-1. Fuente: Elaboración propia....	58
Tabla 35: Id piso, id host, idioma y año de la reseña. Fuente: Elaboración propia.....	58
Tabla 36: Cantidad de comentarios por idioma y su porcentaje agrupado por idioma. Fuente: Elaboración propia.....	59
Tabla 37: Cantidad de reseñas agrupado por año. Fuente: Elaboración propia	59
Tabla 38: Comisión aplicada dependiendo de los valores medios de limpieza . Fuente: Elaboración propia.....	63

1. Introducción

Estamos en un mundo en el que cada vez va todo más rápido, un mundo en el que hemos visto como con el tiempo, las empresas han necesitado tener mayor rapidez en la toma de decisiones para no quedarse atrás. Esta rapidez hay veces que no puede darse debido al gran volumen de datos que gestionan y a la ineficiencia del tratamiento de estos datos.

Desde siempre se han analizado datos y se han utilizado para un fin, ya sea en el ámbito empresarial como en el día a día. Ya lo dijo en el siglo XIX el profesor Kilpatrick, en su artículo “El método de proyecto de Kilpatrick”, (1910) “que los participantes fortalezcan sus competencias docentes, a través del análisis de información específica para planear, desarrollar, presentar resultados y evaluar un proyecto exploratorio”. Antiguamente se analizaban los datos uno a uno, lo que resultaba que fuese una tarea muy poco productiva. Hoy en día, se utilizan softwares, ya sea con una consola de comandos o un programa con interfaz gráfica.

Según *Las 10 mejores herramientas big data*, (2021) las 5 mejores herramientas para desarrollar *data science* son:

1. Lenguaje de programación Python: Es uno de los lenguajes de programación más extendidos gracias a su versatilidad. Esto se debe a que gracias a Python se pueden elaborar una gran cantidad de tareas en muchos sectores distintos. Es muy utilizado en el ámbito del *data science* debido a que es un software con una usabilidad muy sencilla, si lo comparamos con otros lenguajes de programación. Además de todo esto, Python es de código abierto, aspecto que hace que todo el mundo que lo desee pueda colaborar en la mejora de dicho *software*. En contrapartida hay que destacar que tiene una mayor lentitud en comparación con otros *softwares* que están desarrollados para el análisis y gestión de una gran cantidad de datos.
2. Lenguaje de programación R: Se trata de un *software* especializado en el cálculo estadístico. Es la herramienta de *big data* más utilizada. Al igual que Python se trata de un *software* de código abierto, aunque como desventaja señalar que es difícil de utilizar.
3. Hadoop: Se trata de otra de las herramientas de *big data* más importantes. Hadoop tiene, al igual que Python y R, la licencia de código abierto, por lo que hace que sea muy barato utilizarlo y es el *framework* por excelencia para almacenar un gran número de datos. Como desventaja hay que destacar que es complicado de utilizar, por lo que es una barrera de entrada para nuevos usuarios.
4. Apache spark: Se trata de un motor de procesamiento y es uno de los más potentes. Al igual que los 3 anteriores (Python, R y Hadoop) tiene librerías de código abierto, por lo que Apache spark también tiene una constante mejora gracias a la comunidad que va compartiendo sus actualizaciones. Una de las principales ventajas de Apache Spark es que se pueden utilizar diferentes lenguajes de programación como los vistos anteriormente, Python o R, pero también se puede utilizar Java o Scala. A pesar de todo esto, la ventaja más importante es que es hasta 100 veces más rápido que Hadoop.
5. MongoDB: Se trata de una base de datos, pero su principal objetivo es el tratado de documentos.

Gracias al *big data*, cualquier empresa que tenga los conocimientos necesarios puede gestionar una cantidad de datos mucho mayor, y de esa manera, ser mucho más eficiente. Según *Data*

analytics para principiantes, (2020) hay una serie de pasos que normalmente se realizan cuando se hace uso del análisis de datos:

1. Todo empieza con unas preguntas, ya que estas van a ser las que nos indiquen qué datos tenemos que buscar y el rumbo que tenemos que tomar. Un ejemplo de las preguntas que nos haríamos con relación a un caso de estudio de una web que gestiona reservas de alojamiento podría ser: ¿La mayoría de nuestros clientes que realizan una reseña negativa vuelven a reservar con nosotros? O si un cliente realiza una reseña positiva, ¿la siguiente reserva la realizará en el mismo apartamento? Estas son una serie de preguntas que nos harían ver que tenemos que buscar datos relacionados con las reseñas y con los huéspedes.
2. Recopilación de datos: Tras saber qué datos buscar solamente se tiene que ir a los datos internos de la empresa, aunque también se puede buscar datos de otras compañías que sean públicos. De esta manera, podríamos analizar a la competencia y ver posibles ventajas competitivas que tengan en comparación con nuestra entidad. En relación con la recopilación de datos, cabe destacar que tienen que ser datos fiables, ya que, si no, podría ser que estuviésemos haciendo un análisis erróneo tanto de nuestra situación como de la competencia.
3. Organización y procesamiento de los datos.
 - a. Organización de los datos: Es muy importante que el paso siguiente a la obtención de los datos sea organizar los datos y detectar posibles inconsistencias.
 - b. Procesamiento de los datos: Una vez organizados, se debe proceder a realizar la limpieza y estandarización ya que sucede, en numerosas ocasiones, que o bien faltan datos, o bien dichos datos no corresponden con el resto de la columna.
4. Análisis de datos: después de procesar los datos, se puede empezar a analizar los datos. Este análisis no debe tener necesariamente una complejidad extrema. Todo dependerá de la necesidad de información para la toma de decisiones, es decir, de las preguntas del paso 1.
5. Desarrollo de modelos y algoritmos: En el caso de que el análisis se complique y no veamos los resultados claramente se puede hacer uso de modelos estadísticos y algoritmos capaces de analizar muchas más variables a la vez.
6. Visualización de los datos: Con los datos analizados y con la finalidad de poder comprenderlos mejor se utilizan técnicas de representación visual para poder representar los datos. Estos pueden ser gráficas (de barras, de cajas y bigotes, circular, de series comparadas, de líneas verticales) o histogramas, entre otras.
7. Toma de decisiones: Una vez tenemos los datos limpios, organizados, tratados y representados de manera visual, se puede tomar una decisión. En este punto podemos ver las deficiencias o las ventajas competitivas que se tiene en la compañía.

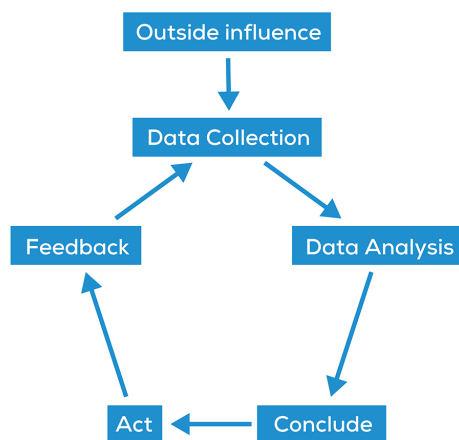


Figura 1. Representación simplificada del ciclo del análisis de datos (Fuente: <https://www.uipath.com>)

El análisis de datos se puede utilizar en todo tipo de sectores, aunque estos son los sectores con mayor aplicación:

- Marketing
- Salud pública
- Deporte
- Investigación
- Seguridad informática
- Mercado inmobiliario

Si bien antes se ha listado los sectores con mayor aplicación, ahora veremos las cuatro áreas en las que se podría aplicar el análisis de datos según *Cuatro áreas en las que puede emplearse el análisis de datos*, (2020):

- Análisis descriptivo: En este caso consistiría en analizar datos recopilados durante un tiempo, para de esta manera tomar decisiones.
- Análisis predictivo: Gracias a este análisis podemos adelantarnos a lo que podría suceder en un futuro. En este caso, y siguiendo con el caso de estudio de una página web de gestión de alojamientos, podríamos intentar averiguar la cantidad de reservas que tendrá la empresa en el próximo año.
- Análisis prescriptivo: En esta área, se analiza y se cuenta cuáles son los pasos a seguir para llegar a un objetivo.
- Diagnóstico analítico: En este caso, se busca encontrar el origen de un suceso para así ayudar a encontrar el motivo de que haya sucedido.

Actualmente, el *data analytics* ha evolucionado hasta tal punto que una gran cantidad de empresas automatizan la obtención, tratamiento y, posterior modelado, de los datos para tener de manera actualizada todos ellos.

Ejemplos de uso en empresas según el artículo: *5 ejemplos de uso real de data analytics*, (2018):

- Uber: Gracias a la localización por Sistema de Posicionamiento Global (GPS) de todos sus vehículos, se logra recopilar una gran cantidad de datos, que le permiten gracias al uso de algoritmos avanzados, optimizar, al máximo, las rutas de recogida y traslado de pasajeros. Por otro lado, Uber también realiza un análisis de la oferta y la demanda que hay, y varía los precios en función de ésta.
- Amazon: Gracias al *data analytics*, la compañía puede crear modelos predictivos y, de esta manera, intentar predecir qué es lo que va a demandar el cliente en un futuro. Por otro lado, Amazon estudia al cliente y le muestra productos que pueden ser de su interés, para que, de esta manera, vea más artículos relacionados con sus gustos y con ello aumente la posibilidad de compra.
- Google Fit y Apple Watch: En este caso, se puede afirmar que el *data analytics* está muy ligado al Internet de las cosas (IoT). En este caso, pequeños dispositivos que, una gran parte de la población lleva encima, captan numerosos datos para su posterior estudio. En el caso de Google Fit y Apple Watch se trata de dispositivos de monitorización de actividad, por lo que, tanto Google como Apple usarán esos datos para ver tendencias de conductas.
- BBVA: En general, toda la banca utiliza el *data analytics*, pero es el banco BBVA el que ha dado un paso más allá y ha sido pionero en el uso de las herramientas en varios proyectos. Con estos proyectos ha logrado obtener información de las actividades comerciales en las grandes zonas urbanas o el impacto que genera (económico) la celebración de un gran evento, como un festival.
- Política: La política juega un papel importantísimo en la sociedad, por lo que tampoco se queda atrás en el uso de las herramientas de *data analytics* para ver las tendencias de la sociedad. La última gran noticia conocida que relacionaba la política y al *data analytics* fue el caso de la empresa de Cambridge Analytica, que supuestamente ayudó para que el presidente Trump saliese electo. Con anterioridad a esto, Obama también hizo uso del análisis de datos para conseguir la reelección en el año 2012, contratando a más de 20 personas encargadas de recopilar e interpretar datos de la sociedad.

2. Objetivos y metodología

El principal objetivo del presente Trabajo Fin de Grado (TFG) es la utilización del análisis de datos para tomar decisiones que permitan obtener una rápida visualización de todos los datos. Con la visualización de datos histórica se identificarán patrones de una manera más sencilla y de ahí ver las tendencias que puede seguir un activo, de forma que se mejore y optimice el proceso de toma de decisiones en cualquier ámbito. Este objetivo se focaliza, específicamente, en analizar todos los aspectos relacionados con un caso de estudio, particularmente con la compañía Airbnb desde el punto de vista de las reservas. Airbnb es una compañía relativamente nueva pero que ha sabido, gracias a su forma de negocio de economía colaborativa, convertirse en una de las mejores compañías de la industria de la gestión de reservas de alojamiento.

En cuanto a los objetivos específicos se detallan a continuación:

1. Buscar datos relevantes que sirvan para entender cómo son las reservas en Airbnb, tanto desde el punto de vista del cliente, como desde el punto de vista del anfitrión (*host*). Estos datos tienen que ser reales y acordes con lo que se busca, de ahí su dificultad de encontrarlos. En este caso, se desea encontrar los datos de la ciudad de origen objetivo que se ha considerado: Valencia.
2. Limpiar datos para la correcta manipulación de estos.
3. Crear las diferentes tablas para un análisis más sencillo.
4. Crear las agrupaciones necesarias para entender mejor las tablas.
5. Representar gráficamente de los datos obtenidos para la correcta visualización y para una mayor facilidad de análisis.
6. Extraer y analizar las conclusiones del conjunto de datos masivos.

La metodología que se ha utilizado en el presente TFG se muestra en la Figura 2.



Figura 2. Metodología del TFG (Fuente: Elaboración propia)

En la figura 2 se encuentra simplificada la metodología.

Por un lado tenemos las fuentes de información, donde se muestra como se han tratado los datos, y a que problemas se ha tenido que hacer frente. Entre ellos encontramos la limpieza de datos, y la nueva incorporación de otros como es el género o el idioma de las reseñas

Por otro lado, en los pasos metodológicos se puede apreciar cada uno de los pasos seguidos en este trabajo, siendo en primer lugar el estudio de la importancia del *data analytics* en la toma de decisiones, seguido por la selección del caso de estudio con la obtención y limpieza de los datos. Seguido de esto se realizó un exhaustivo análisis de los



mismos, y una vez se tuvo los datos analizados se comprobó que había partes del proceso mejorables y se comentó las formas de mejorar dichos procesos.

Finalmente, en la columna de resultados, tenemos la parte de lo que ha dado el trabajo. Esto ha sido en primer lugar, gracias a una herramienta poco usada por las empresas, que nos ha dado unos datos que han permitido ver las posibles mejoras que pueden afectar a los barrios, por géneros, por zonas, por reseñas y por el idioma de la persona que reserva.

3. Caso de Estudio

3.1. Airbnb

3.1.1 Historia

Airbnb nació en el año 2008 en San Francisco después de que sus 2 creadores encontraran una oportunidad de negocio en el mercado de los alquileres de pisos por días. Los dos creadores empezaron alquilando su propia casa debido a que en la ciudad donde residían había un evento y se agotaron las habitaciones de hotel disponibles. Al ver esta oportunidad, crearon rápidamente una página web que les permitió mostrarse a mucha más gente.

El nombre de la compañía proviene de *air* junto con las letras 'Bnb', que proviene de '*Bed and Breakfast*'.

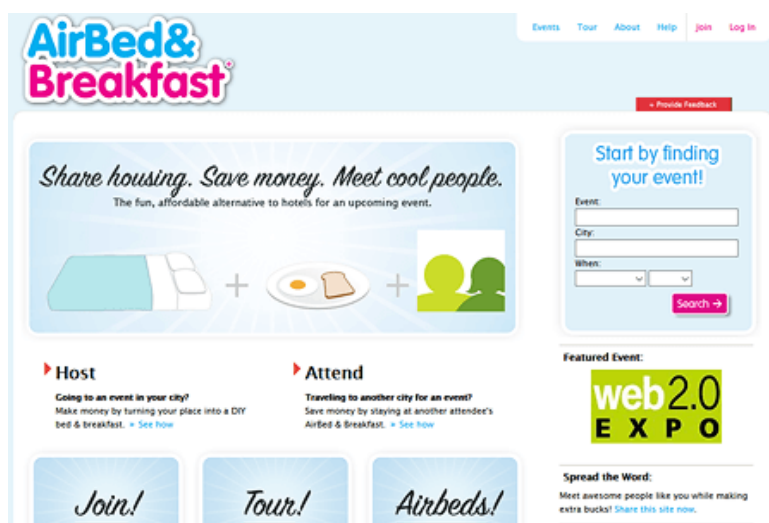


Figura 3. Airbnb en el año 2010 (Fuente: www.webdesignmuseum.org)

Los primeros problemas vinieron cuando empezaron a expandir su negocio y se encontraron con problemas con los inversores, que les transmitían que no tendrían futuro, ya que el modelo de negocio consistía en dejar entrar a tu casa a un desconocido.

Para poder hacer frente a este problema, la compañía tuvo que hacer diferentes esfuerzos para conseguir la confianza de la gente y, por encima de todo, de los inversores. Esto, en parte lo consiguieron gracias a la creación de una ficha pública que tiene cada usuario que se registra en Airbnb. En esta ficha aparecen tanto las reseñas que le han puesto a cada usuario como si está o no verificado por la compañía.

3.1.2 La empresa

Hoy en día, Airbnb ofrece un mercado comunitario, tanto de apartamentos como de experiencias, dónde, de manera muy económica, se puede listar una propiedad y, de esta manera, tenga una visibilidad prácticamente mundial (190 países). Todo esto hace que sea uno de los casos con mayor éxito de la economía colaborativa. Airbnb tiene presencia en 165.000 ciudades repartidas por todo el mundo y presenta más de 4 millones de alquileres en total.

Por otro lado, al ser la empresa Airbnb la que gestiona cada reserva permite tanto al *host* (propietario del inmueble), como al inquilino tener una tranquilidad extra. Esto se debe a que, si hay algún problema, Airbnb mediará para solucionar el incidente. Otra cosa más que la compañía ofrece es un servicio de atención al cliente 24 horas, 7 días a la semana.

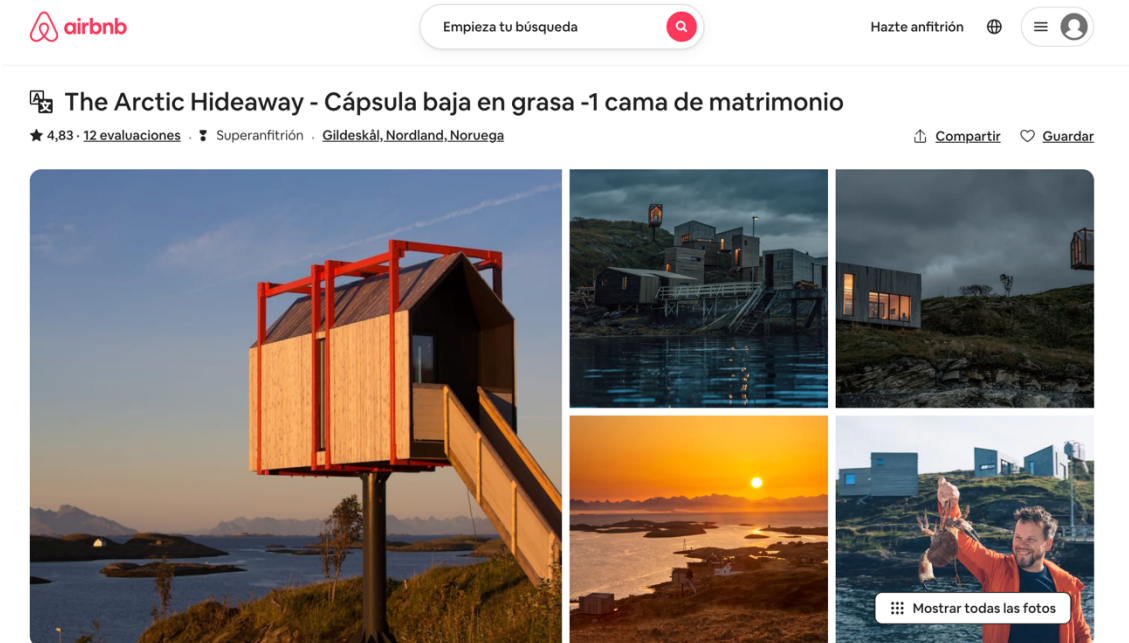


Figura 4. Habitación privada en una isla (Fuente: www.airbnb.es)

Para poder utilizar el servicio, primero es necesario crear una cuenta, ya que va a ser la que permita a los usuarios realizar la reserva. Una vez creada la cuenta, se puede buscar de diferentes maneras, utilizando los filtros generales hasta con el uso de filtros más específicos como filtrar por habitación, cantidad de camas, baños, tipo de propiedad, si tiene o no servicios básicos como wifi, secador, cocina, aparcamiento... y otros muchos más. En total existen más de 25 filtros.

3.1.3 Motivación de la elección de Airbnb

Tal y como se ha comentado anteriormente, Airbnb maneja una gran cantidad de datos. Estos datos no solo son la cantidad de alquileres que hay, sino todo lo que rodea a estos alquileres, es decir, desde los precios hasta las reseñas, pasando por las características del apartamento (camas, habitaciones, disponibilidades pasadas...).

Es por todo ello, y porque se trata de una compañía que es interesante analizar cómo funciona, por lo que se ha escogido Airbnb.

Finalmente, también cabe destacar que parte de la decisión de escoger Airbnb ha sido debido a que el turismo en España es muy importante y siempre es positivo para la sociedad ayudar a detectar errores a las empresas que hacen posible que el turismo crezca.

3.2. Obtención del conjunto de datos

Todos los datos que se han utilizado han sido de terceros. Por un lado, todos los datos de Airbnb han sido sacados de la web (Get the Data, 2022) <http://insideairbnb.com/get-the-data>. En dicha web encontramos diferentes tablas de las que se ha obtenido información muy valiosa:

1. Calendar.csv.gz: Muestra si estará disponible cada uno de los apartamentos listados en Valencia durante todo lo que resta de año 2022, así como los precios que tendría reservarlo dichos días y el número mínimo y máximo de noches que se puede reservar (Tamaño de la tabla: 2199125 filas X 6 columnas).
 - a. Listing_id: Identificador. Valor único que se da a cada apartamento en Airbnb.
 - b. Date: Fecha del día en el que se comprueba el precio y disponibilidad.
 - c. Available: Valor booleano. Indica 't' si está disponible (true) y 'f' si no lo está (false).
 - d. Price: Precio del alojamiento para ese día.
 - e. Minimum_nights: Noches mínimas que se puede alquilar por cada reserva dicho apartamento.
 - f. Maximum_nights: Noches máximas que está en alquiler por cada reserva dicho apartamento.
2. Listings.csv.gz: Se trata de la tabla con información relevante. Se va a hacer un pequeño resumen de las columnas más significativas: (Tamaño de la tabla: 6025 filas X 74 columnas)
 - a. Room_type: Muestra el tipo de propiedad. En este caso hemos encontrado solamente 4: casas enteras, habitaciones compartidas, habitaciones privadas y habitaciones de hotel.
 - b. De la parte de las características de la propiedad podemos encontrarnos "baños", "camas", "habitaciones", y "precio".
 - c. De la parte de las reseñas que tiene dicha propiedad podemos encontrarnos los 7 tipos de reseñas que tiene un apartamento de Airbnb: "Reseña de limpieza", "Reseña de comunicación" y "Reseña general", entre otras.
 - d. De la parte del *host*, podemos encontrarnos desde la cantidad de propiedades que posee en Airbnb y el tipo de cada una de ellas, hasta desde cuando está registrado en Airbnb, si es *superhost* o no, y los ratios de aceptación y de respuesta... entre otros.
3. Neighbourhoods.csv: Esta tabla muestra los barrios de la ciudad. En este caso, se agrupan en barrios y sub-barrios. (Tamaño de la tabla: 88 filas X 2 columnas):
 - a. Neighbourhoods_group: Muestra el nombre del barrio. Hay 19 diferentes.
 - b. Neighbourhood: Muestra el nombre del sub-barrio. Hay 88 barrios diferentes.
4. Neighbourhoods.geojson: Esta tabla muestra la información relativa a la geoposición de cada uno de los barrios. Posteriormente, se utilizará para indicar en qué zona del mapa están dichos barrios.

5. `Reviews.csv.gz`: Esta tabla muestra todas y cada una de las reseñas puestas en Airbnb en Valencia. Aparece tanto el id del huésped, como el día de la reseña (Tamaño de la tabla: 244812 filas X 7 columnas).

Por otro lado, el resto de información de los barrios se ha obtenido en la web del Ayuntamiento de Valencia <https://www.valencia.es/val/estadistica/mapa-districtes>. Estos datos han permitido obtener una cantidad de información mayor.

3.3. Procesado y limpieza de datos

3.3.1. Problemas con la limpieza de los datos

Para hacer operaciones con los datos es necesario que estén en el mismo formato y que, al mismo tiempo, tengan un formato operable, es decir, no se pueden sumar dos valores alfanuméricos. En Python existen, entre otros muchos formatos, los valores enteros (int) y los valores *string* o cadenas (str). Si bien hubiera una columna con valores *string* habría que pasarlos a enteros.

En este caso, se debe realizar un bucle *for* que recorra toda la tabla y cambie los valores de la columna deseada. Para ello, y a modo de resumen, dentro del bucle *for* se debería introducir el carácter 'int(' seguido del nombre de la tabla junto con el nombre de la columna y finalmente cerrando el paréntesis ')'. De esta manera se conseguiría tener la columna con valores enteros y se podría empezar a trabajar con esos datos.

En cuanto a las filas con valores *NaN* o valores nulos se puede optar por varias opciones para limpiar los datos:

- Eliminar todas las filas que presenten algún valor *NaN* en cualquiera de las columnas. Esto es recomendable como solución rápida si hay pocas casillas con valores *NaN*. Abusar de esto puede provocar que la tabla se quede muy pequeña y se pierda mucha información. Esto se realiza mediante el siguiente comando, siendo *df* la variable de la tabla y todo aquello que va entre " el nombre de la columna:
 - `df = df[df['comments'].notna()]`
 - `df.reset_index(drop=True, inplace=True)`
- Transformar todo valor *NaN* a 0. Este procedimiento está apropiado siempre y cuando se tenga muchos valores *NaN* y no se desee eliminar las filas. El problema de esto es que cabe la posibilidad que contando los valores '0' estemos obteniendo resultados distorsionados.
 - `df = df.fillna(0)`

En lo que se refiere a la columna de precio, situada en la tabla 'listings.csv.gz' se ha podido observar que el dato posee el símbolo de la divisa, por lo que ha habido que intervenir para quitarlo. Si bien es cierto que esto es una cadena de texto, no se puede transformar en entero con lo explicado anteriormente ya que daría error por el símbolo. En primer lugar, hay que borrar el símbolo, mediante un bucle *for*, recorriendo toda la tabla buscaremos de cada celda, el valor deseado y se modificará por hueco en blanco.

- `for dias in df.iterrows():`
 - `index_value, column_values=dias`

- `sindolar=str(column_values['price'].replace('$',''))`
- `toint=pd.to_numeric(sindolar, downcast='integer')`
- `df.at[index_value,'price OK']=toint`

Finalmente, en la columna de baños se encontró un problema debido a que había una gran cantidad de datos distintos, por lo que se tuvo que hacer algo similar al apartado anterior con el símbolo de la divisa.

- `for dias in df.iterrows():`
 - `index_value, column_values=dias`
 - `arreglo=str(column_values['bathrooms_text']).replace(' shared bath','')`
 - `arreglo2=arreglo.replace(' bath','')`
 - `arreglo3=arreglo2.replace(' baths','')`
 - `arreglo4=arreglo3.replace(' private','')`
 - `arreglo5=arreglo4.replace('s','')`
 - `arreglo6=arreglo5.replace('Private half-bath','')`
 - `arreglo7=arreglo6.replace('nan','')`
 - `df.at[index_value,'bathrooms_text OK']=arreglo7`

3.3.2 Procesado de datos

Una vez limpiados los datos, se analiza qué datos extras hacen falta. En este paso se han realizado principalmente 2 tareas:

- Agregación de una nueva columna llamada “Género”. Los datos de esta columna se han tomado a partir de los datos de la columna “host_name”, es decir nombre del *host*. Se ha podido llevar a cabo la clasificación mediante técnicas de comparación. Esta clasificación ha sido entre hombres, mujeres y grupos, ya que también hay hoteles, inmobiliarias o grupos de personas.
 - `for dias in df.iterrows():`
 - `index_value, column_values=dias`
 - `if(column_values['host_name'][-1:]=='a' or column_values['host_name'][-2:]=='en' or column_values['host_name'][-3:]=='har' or column_values['host_name'][-3:]=='bel' or column_values['host_name'][-2:]=='es' or column_values['host_name'][-1:]=='Maria Jos' or column_values['host_name'][-3:]=='ion' or column_values['host_name'][-2:]=='iz' or column_values['host_name'][-2:]=='ne' or column_values['host_name'][-3:]=='cio' or column_values['host_name'][-3:]=='rat'`
 - ...
 - `df.at[index_value,'Gen']='Mujer'`
 - `else:`

- df.at[index_value,'Gen']='Hombre'
 - if(column_values['host_name'][1:]=='afa' or
column_values['host_name'][1:]=='oises' or
column_values['host_name'][1:]=='gnacio' or
column_values['host_name'][1:]=='avier'):
 - df.at[index_value,'Gen']='Hombre'
 - if(' y ' in column_values['host_name'] or ' Y ' in column_values['host_name']
or '&' in column_values['host_name'] or '+' in column_values['host_name']
or '4' in column_values['host_name'] or '5' in column_values['host_name']
or '7' in column_values['host_name'] or 'SL' in column_values['host_name']
or 'S.L' in column_values['host_name']
 - ...
 - df.at[index_value,'Gen']='Grupo'
- Clasificación del idioma de las reseñas para su posterior tratamiento. En este caso, no se ha hecho una comparación masiva de la columna respecto a una cantidad grande de datos, sino que se ha utilizado un módulo que ha permitido conocer el idioma del texto con solo pasarle 2 parámetros.
 - for i in df.index:
 - df.at[i,'comments']=langid.classify(df['comments'][i])[0]
 - fechapr=df['date'][i][0:4]
 - fechapr=int(fechapr)
 - df.at[i,'año']=fechapr

3.4. Análisis de datos

Una vez terminado el paso anterior, y con los datos necesarios para realizar el análisis, se procede a ello. En este caso, se ha dividido en diferentes secciones que tienen en común la agrupación de los datos. A pesar de que se haya hecho la división no implica que no se usen datos de un apartado en otro.

3.4.1 Tipo de alquileres

En este apartado se va a realizar un análisis sobre el tipo de alquileres que hay en Airbnb.

En primer lugar, se obtendrá información sobre cada uno de los distintos tipos de alquileres que hay, tal y como se puede observar en la Tabla 1.

Tabla 1: Cantidad total de alquileres por tipo. Fuente: Elaboración propia

Tipos pisos/habitaciones por tipo	Cantidad
Entire home/apt	4143
Private room	1827
Shared room	36
Hotel room	19

Tal y como se puede observar, la gran mayoría de opciones en Airbnb son casas enteras, siendo más del doble que la segunda opción que son habitaciones privadas. Por otro lado, en cuanto a las habitaciones compartidas y a las habitaciones de hotel es un número muy reducido. Los porcentajes respecto al total son:

Tabla 2: porcentaje total de alquileres por tipo. Fuente: Elaboración propia

Tipos pisos/habitaciones por tipo	Porcentaje
Entire home/apt	68,76%
Private room	30,32%
Shared room	0,59%
Hotel room	0,31%

A continuación, se ha clasificado los tipos de alquileres y por sub-barrios, obteniendo así una tabla de 88 filas x 5 columnas (Tabla 3).

Tabla 3: Cantidad de alquileres de cada tipo en cada barrio. Fuente: Elaboración propia

	Neighbourhood	Geometry	Q de casas enteras	Q hab. priv	Q hab. comp
0	BENIFARAIG	MULTIPOLYGON (((-0.37639 39.52831, -0.37667 39...	0.0	1.0	0.0
1	BENICALAP	MULTIPOLYGON (((-0.38139 39.49869, -0.38120 39...	56.0	35.0	1.0
2	TORREFIEL	MULTIPOLYGON (((-0.37122 39.49839, -0.37180 39...	16.0	18.0	0.0
3	TORMOS	MULTIPOLYGON (((-0.38216 39.49029, -0.38050 39...	20.0	14.0	0.0
4	SANT ANTONI	MULTIPOLYGON (((-0.37388 39.49206, -0.37273 39...	12.0	14.0	0.0
...
86	EL FORN D'ALCEDO	MULTIPOLYGON (((-0.38868 39.43346, -0.38868 39...	2.0	2.0	0.0
87	LA TORRE	MULTIPOLYGON (((-0.39403 39.42720, -0.39402 39...	0.0	0.0	0.0

Las columnas que se han utilizado son:

- Neighbourhood: Enumera el sub-barrio en el que está situada el alojamiento.
- Geometry: Estos datos son imprescindibles para poder geolocalizar el alojamiento e imprimir en el mapa los datos de cada barrio.
- Cantidad de casas enteras.
- Cantidad de habitaciones privadas.
- Cantidad de habitaciones compartidas.

Como se ha comentado, gracias al uso de la columna "Geometry" se puede localizar en un mapa, la cantidad de pisos que hay de cada tipo de manera muy visual.

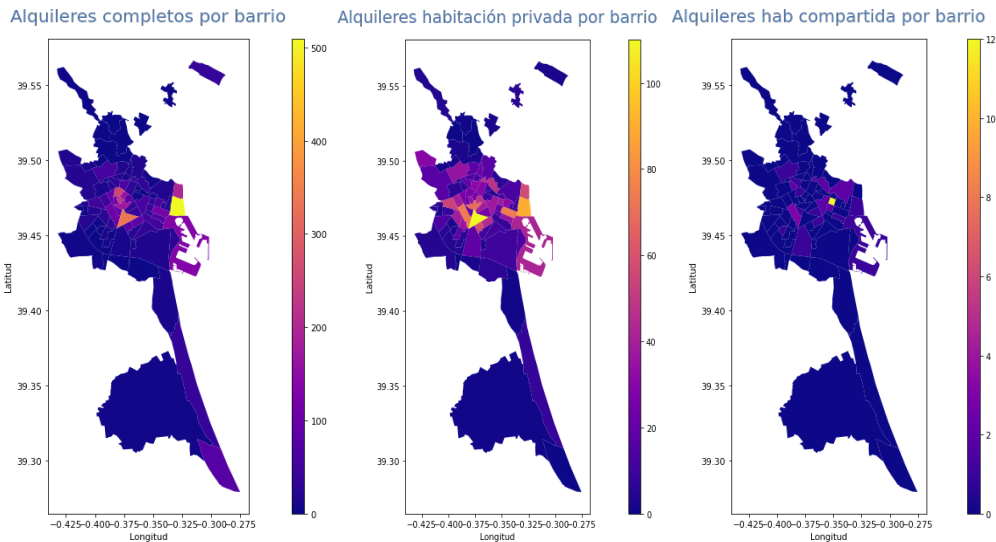


Figura 5. Representación por sub-barríos de cantidad de cada tipo de alquileres (Fuente: Elaboración propia)

Se puede observar que tanto en las casas completas como en las habitaciones privadas hay una gran concentración de pisos en el centro de la ciudad y en la zona de la playa de Valencia .

Teniendo en cuenta los pisos enteros, los barrios con mayor cantidad de oferta son:

- CABANYAL-CANYAMELAR
- RUSSAFA
- EL CARME
- EL MERCAT
- LA MALVA-ROSA

Todos estos barrios presentan entre 509 y 200 pisos completos.

Teniendo en cuenta las habitaciones privadas, los barrios con mayor cantidad de oferta son:

- RUSSAFA
- CABANYAL-CANYAMELAR
- AIORA
- ARRANCAPINS
- SAN FRANCESC

Todos estos barrios presentan entre 110 y 69 habitaciones privadas cada uno.

Teniendo en cuenta las habitaciones compartidas, los barrios con mayor cantidad de oferta son:

- L'AMISTAT
- ARRANCAPINS
- MESTALLA
- BENIMACLET
- ALBORS

Todos estos barrios presentan entre 12 y 2 habitaciones compartidas cada uno.

Una vez representados visualmente cómo se distribuyen los tipos de alquileres por barrios, se procede con el análisis de los tipos de alquileres. En este caso se ha agrupado en una tabla cada tipo de alquiler y se ha considerado las siguientes columnas para su posterior estudio:

- Numero de reseñas: Muestra en número, la cantidad de reseñas que tiene cada alojamiento.
- Reseñas: Muestra, del 1 al 5, el valor medio de cada tipo de reseña por cada alojamiento.
 - Reseña de precisión.
 - Reseña de limpieza.
 - Reseña de check-in.
 - Reseña de comunicación.
 - Reseña de localización.
 - Reseña general.
- Precio: Muestra el precio en el momento de la recogida de los datos. En este caso el día de la recopilación de datos fue el 30 de marzo del 2021.

Tabla 4: Cantidad de reseñas y sus valores por tipo de alquiler. Fuente: Elaboración Propia

	Tipo de alquiler	Num reseñas	Reseña precisión	Reseña Limpieza	Reseña Check-in	Reseña Comunicación	Reseña Localización	Reseña General	€
0	Entire home/apt	120	4.64	4.58	4.82	4.73	4.42	4.50	66.0
1	Entire home/apt	1	5.00	5.00	5.00	5.00	5.00	5.00	100.0
2	Entire home/apt	103	4.73	4.89	4.85	4.75	4.75	4.67	142.0
3	Entire home/apt	133	4.81	4.58	4.86	4.89	4.80	4.65	107.0
...
5104	Private room	0	NaN	NaN	NaN	NaN	NaN	NaN	47.0

En este caso no se han eliminado las columnas con *NaN* porque corresponden a los alquileres que no han tenido todavía reseñas por lo que es importante mantenerlas para su posterior estudio.

Para poder visualizar los datos todavía con mayor claridad, se procede a hacer una agrupación por tipo de alquiler, considerando las columnas “número de reseñas”, “valor medio de reseñas”.

Tabla 5: Valor reseña global y cantidad media de reseñas por tipo de apartamento. Fuente: Elaboración propia

Tipo de alquiler	Cantidad alquileres	Número de reseñas	Valor reseña global	Reseñas medias/alquiler
Entire home/apt	4143	194797	4.611491	47.018344
Hotel room	19	339	4.545000	17.842105
Private room	1827	49612	4.568183	27.154899
Shared room	36	73	4.610500	2.027778

Viendo los resultados de la Tabla 5, se observa que los tipos de alquiler que tienen una media de mayor cantidad de reseñas son las casas enteras, seguidos por las habitaciones privadas. Teniendo en cuenta los valores medios de las reseñas, se observa que también son superiores la valoración de las casas enteras pero esta vez le sigue las habitaciones compartidas.

Estos datos dan a entender que las personas que reservan en Airbnb prefieren alquilar una casa entera y que, después de dejar el apartamento, están más satisfechos.

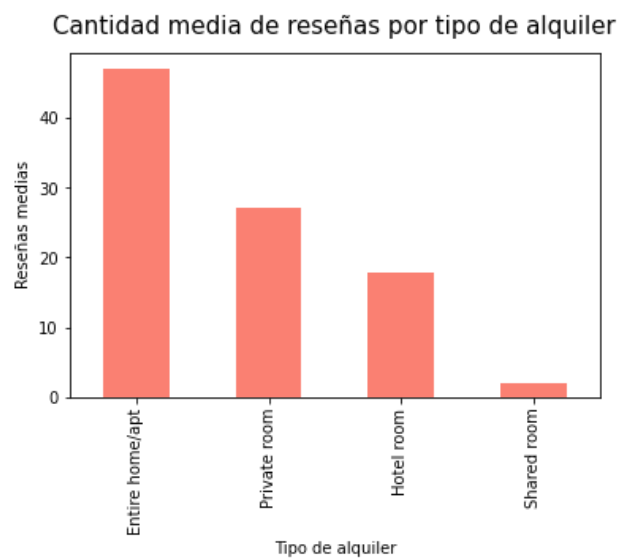


Figura 6. Cantidad media de reseñas por tipo de alquiler (Fuente: Elaboración propia)

Una vez identificadas las valoraciones medias, también sería interesante indagar en cada una de las distintas valoraciones que tienen los alquileres. Para ello se procede a agrupar por tipo de alquiler.

Tabla 6: Valor media de todas las reseñas por tipo de alquiler. Fuente: Elaboración propia

	Cantidad alquileres	Reseña precisión	Reseña limpieza	Reseña Check-in	Reseña comunicación	Reseña localización
Entire home/apt	4143	4.725712	4.628225	4.799959	4.795160	4.721215
Hotel room	19	4.650000	4.678333	4.758889	4.652778	4.626667

	Cantidad alquileres	Reseña precisión	Reseña limpieza	Reseña Check-in	Reseña comunicación	Reseña localización
Private room	1827	4.688457	4.588065	4.776213	4.738191	4.695330
Shared room	36	4.637500	4.640625	4.751875	4.711250	4.760625

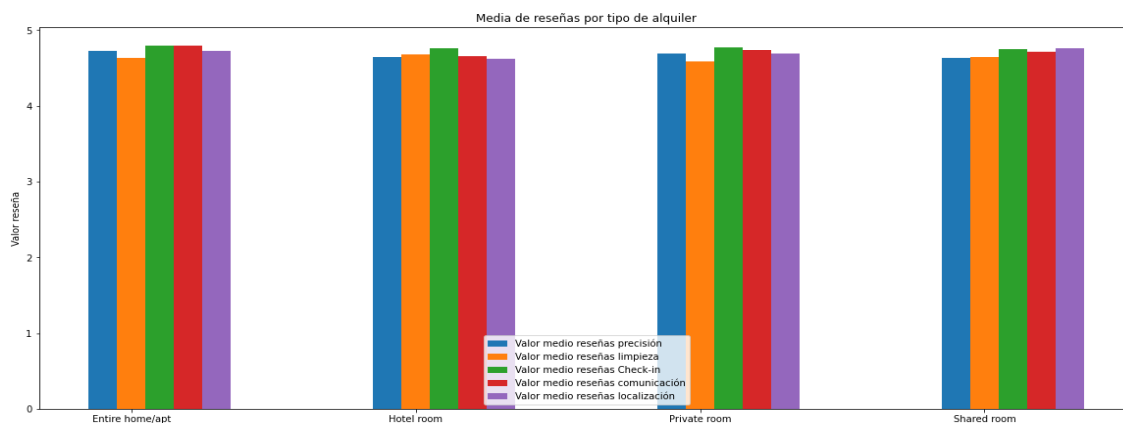


Figura 7. Representación de la media de todas las reseñas por tipo de alquiler (Fuente: Elaboración propia)

Como se puede observar en la figura anterior, las reseñas medias de todas las opciones de alquiler se sitúa entre 4 y 5, aspecto que es muy positivo, ya que indica que los inquilinos normalmente valoran de forma muy positiva la estancia. Aun así, cada alquiler tiene unas reseñas mejores y otras peores:

- Casa entera: Las reseñas en las que más destacan de forma positiva los alquileres de casas enteras es en el Check-in y en la comunicación *host*-inquilino, mientras que las peores reseñas en estos tipos de alquiler son en la limpieza.
- Habitación de hotel: Las reseñas que más destacan de forma positiva en las habitaciones de hotel es, como es de esperar, en el Check-in, ya que en este tipo de alquileres hay una recepción (muchas veces 24 h) y, de manera secundaria, en la limpieza. Las peores reseñas de estos tipos de alquiler se encuentran en la localización, siendo lógico debido a que hay menos cantidad de hoteles que de apartamentos por lo que pueden estar en menos zonas.
- Habitación privada: Las reseñas que más destacan de forma positiva en las habitaciones privadas es en el Check-in, y de manera secundaria, en la comunicación *host*-inquilino. Las peores reseñas de estos tipos de alquiler se encuentran en la limpieza.
- Habitación compartida: Las reseñas que más destacan de forma positiva en las habitaciones compartidas es en la localización, mientras que las peores reseñas de estos tipos de alquiler se encuentran en la precisión.

Estos últimos datos son medias, por lo que no implica que no haya datos que estén por encima y por debajo de esas medias, como por ejemplo, cualquier reserva de habitación de hotel vaya a tener una mejor reseña de Check-in que el resto.

A partir de la tabla anterior, se realizará una correlación para ver qué variables están relacionadas entre sí de manera positiva o negativa y cuáles no están relacionadas entre sí.

En Python hay una manera sencilla de analizar la correlación que tienen entre las variables de una tabla, usando el comando "corr()".

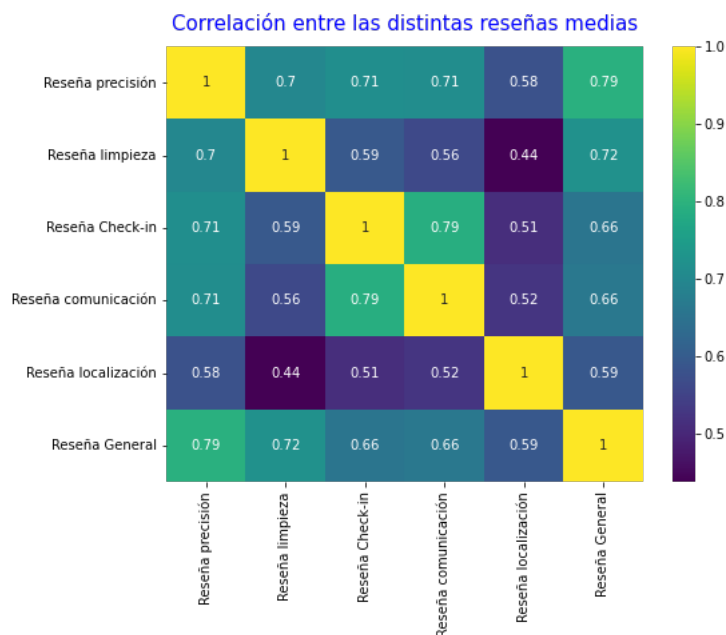


Figura 8. Correlación entre las distintas reseñas (Fuente: Elaboración propia)

Toda aquella correlación de dos variables que se sitúe entre -0,7 y 0,7 querrá decir que están poco o nada correlacionadas. A partir de 0,7 hasta 1 significa que están muy relacionadas positivamente, es decir, que cuando una variable aumenta, la otra también. Por el otro lado, si el valor se encuentra entre -0,7 y -1 quiere decir que están muy relacionadas negativamente, es decir, cuando una variable aumenta, la otra disminuye, y por el contrario, cuando una disminuye, la otra aumenta.

Gracias a la representación de la correlación se puede observar cuales son las variables que más están correlacionadas entre sí.

- Reseña de precisión-Reseña de limpieza: 0,7
- Reseña de precisión-Reseña Check-In: 0,71
- Reseña de precisión-Reseña de Comunicación: 0,71
- Reseña de precisión-Reseña General: 0,79
- Reseña de Limpieza-Reseña General: 0,72
- Reseña Check-In-Reseña de Comunicación: 0,79

Teniendo en cuenta estas correlaciones positivas, cabe destacar que normalmente quien tiene una reseña de precisión buena suele tener buenas reseñas en limpieza, check-in, comunicación y en la reseña general, es decir, en todas, menos en la localización.

3.4.2 Hosts

Los usuarios que alquilan las casas (*hosts*) o la propia plataforma Airbnb son piezas muy importantes para que todo funcione correctamente, pero sobre todo, uno de los activos más importante son los *hosts*. Estas, son personas que ceden sus casas o habitaciones individuales a cambio de una compensación económica. Al dejarle la casa a un tercero está arriesgando su bien. Es por ello que hay que estudiar qué requisitos deben cumplir los *hosts*, para que Airbnb premie a los que mejor lo realizan y ayude a mejorar a los que menos rendimiento tengan. De esta manera se conseguirá tener a los *hosts* contentos y Airbnb podrá seguir funcionando.

En primer lugar, se estudia cuántas casas tiene cada *host*. Esto se puede realizar con relativa facilidad debido a que se tiene indexado el identificador (*id*) de cada *host* en cada uno de los alquileres de Airbnb.

Para conocer el número total de alquileres que tiene cada *host* se realizará una agrupación por cada *id host* y se sumará la cantidad de pisos o habitaciones listadas en Airbnb.

- `top_host=df.host_id.value_counts()`
- `top_host.head(15)`

Los resultados obtenidos se muestran en la Tabla 7 que incluye el listado del *id* del *host* junto con la cantidad de alquileres que tiene listados.

Tabla 7: Valor media de todas las reseñas por tipo de alquiler. Fuente: Elaboración propia

Host_id	Q alquileres
83066665	139
409088222	37
14182776	37
97229818	36
134051803	34
123160301	30
357999866	29
427643509	28
396941410	28
308967204	24
102196656	23
12044109	22
97743716	22
2037083	20
63053029	20

Como se aprecia el *host_id* "409088222" tiene un total de 37 alquileres, y por lo que respecta a esa cantidad, se debe de tratar de una inmobiliaria o alguna entidad de gestión de pisos.

Este dato se puede comprobar mediante la siguiente búsqueda:

- `df[(df.host_id==409088222)]['listing_url']`

Dicha búsqueda da como resultado el listado de todos y cada uno de los alquileres que tiene en la plataforma dicho *host* (*id* del apartamento o habitación a alquilar junto con la *url* de dicho apartamento o habitación):

4671	https://www.airbnb.com/rooms/50632852
4715	https://www.airbnb.com/rooms/50787565
4756	https://www.airbnb.com/rooms/50890366
4791	https://www.airbnb.com/rooms/51012677
	...
5760	https://www.airbnb.com/rooms/567221750614028537
5761	https://www.airbnb.com/rooms/567228425909298698

Si se accede a en cualquier enlace anterior, se puede comprobar que se trata de un piso en Valencia, y si se accede dentro del perfil, se puede comprobar que es un *host* con 37 alquileres. Cabe destacar que desde que se descargan los datos, hasta que se procesan, puede haber variaciones.

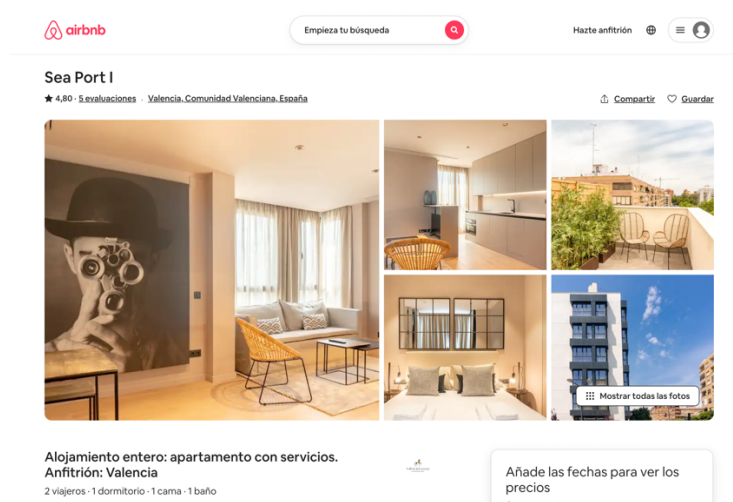


Figura 9. Página Airbnb apartamento con servicios (Fuente: Airbnb.es)

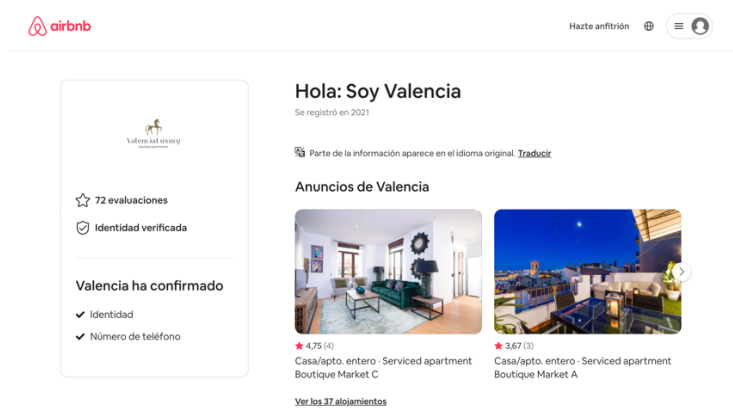


Figura 10. Cuenta de usuario Airbnb (Fuente: Elaboración propia)

Para comprobar el número total de *hosts* que hay se calculará la cantidad de *host_id* que hay distintos y se sumarán.

Para ello, se puede reutilizar parte de la fórmula anterior y hacer un conteo de los distintos:

- `top_host.count()`

En el caso de Airbnb en Valencia, hay 3339 *hosts* distintos.

Siguiendo con la información que se está analizando de los *hosts*, se procede a definir una tabla (Tabla 8) que contenga los *ids* de los *hosts*, la fecha desde cuando están registrados, su ratio de respuesta, su ratio de aceptación, si son *superhost*, y si tienen licencia.

Tabla 8: Ratio de respuesta y de aceptación en función del *superhost* y licencia. Fuente: Elaboración propia

	id	desde	ratio respuesta	ratio aceptación	superhost	licencia
0	219476	2010-09-01	100%	100%	f	VT-41540-V
1	120649	2010-05-08	NaN	NaN	f	NaN
2	508042	2011-04-14	100%	35%	f	31815
3	1451371	2011-11-28	NaN	78%	f	NaN
4	535881	2011-04-26	100%	63%	f	VT-36438-V

Los resultados de la tabla 8 muestran que hay un formato específico en tipo fecha, y para solucionar dicho problema, se necesitará averiguar la cantidad de días que han transcurrido desde ese día hasta la actualidad. Por ello se realizará la siguiente operación:

- for i in range(6025):
 - `varfecha=datetime.strptime(hoster['Desde'][i], '%Y-%m-%d')`
 - `dias=datetime.now()-varfecha`
 - `hoster.at[i,'Desde']=int(dias.days)`
- for e in range(6025):
 - `porcemenos=str(hoster['Ratio Respuesta'][e]).replace('%', '')`
 - `hoster.at[e,'Ratio Respuesta']=porcemenos`
 - `porcemenos=str(hoster['Ratio Aceptacion'][e]).replace('%', '')`
 - `hoster.at[e,'Ratio Aceptacion']=porcemenos`
- `hoster.columns=['id','Desde','Ratio respuesta','ratio aceptación','superhost','licencia']`
- `hoster=hoster.dropna(subset=["Ratio respuesta"])`
- `hoster`

Y de esta forma se obtiene la Tabla 9, con la fecha transformada en días transcurridos (Tabla 9 reducida para la memoria).



Tabla 9 :Ratio de respuesta y de aceptación en función de superhost y licencia con fecha en días. Fuente: Elaboración propia

	id	Desde	Ratio respuesta	ratio aceptación	superhost	licencia
0	219476	4342	100	100	f	VT-41540-V
1	120649	4458	nan	nan	f	NaN
2	508042	4117	100	35	f	31815
3	1451371	3889	nan	78	f	NaN
4	535881	4105	100	63	f	VT-36438-V

Siguiendo con el análisis anterior, se pretende comprobar si existe alguna relación entre el ratio de respuesta y de aceptación junto con la cantidad de días que lleva el alojamiento disponible en la plataforma, ya que preocupa que los usuarios tengan comportamientos distintos al cabo de los años.

Para ello, se va a realizar una agrupación de todos los ratios de respuesta iguales y se va a hacer la media de todos los días medios que llevan en la plataforma los respectivos *hosts*.

Tabla 10: Días medios del host en Airbnb agrupado por ratio de respuesta. Fuente: Elaboración propia

Ratio respuesta	Días medios en la plataforma
99	2163.065089
98	2337.630769
97	2400.775510
96	2304.222222
...	...
14	2127.000000
13	2130.250000
0	1818.750000

La figura 11 muestra la representación gráfica de los datos anteriores, para facilitar la comparativa visual.

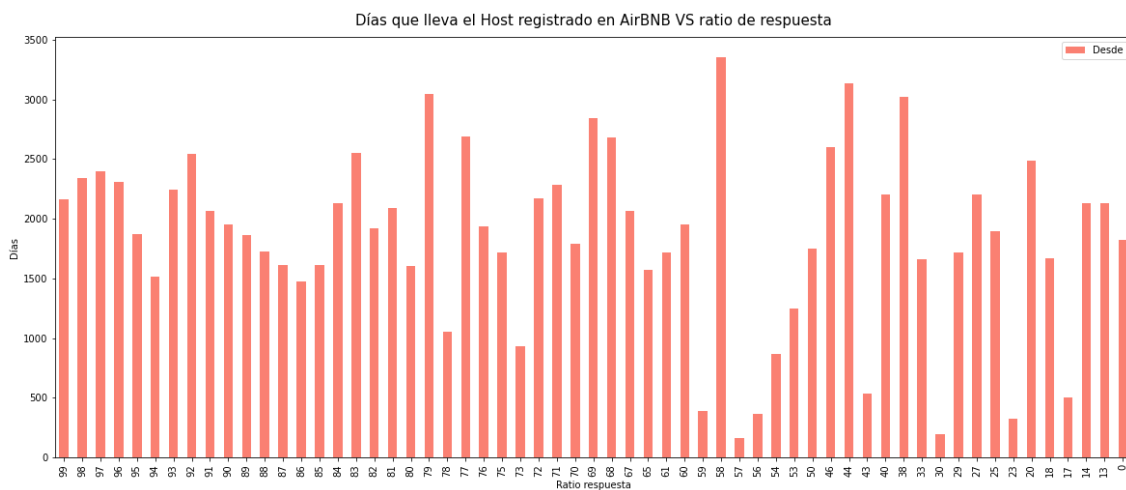


Figura 11. Gráfica de días medios que lleva el host en Airbnb por el ratio de respuesta (Fuente: Elaboración propia)

Como se puede observar, quien presenta un 58% de respuesta es el grupo con mayor tiempo en la plataforma, mientras que el valor inmediatamente inferior, el 57% pertenece a los que menos tiempo tienen en la plataforma. Es por ello, que no existe una correlación entre el tiempo que lleva un usuario en la plataforma y el ratio de respuesta.

A pesar de no haber correlación, sí que cabe destacar la cantidad de *hosts* que hay por encima del 90%. Cabe destacar que, para dar un servicio correcto es importante que se responda con rapidez pero, sobre todo, que se responda.

Tabla 11: Cantidad de host que responden por cada ratio de respuesta. Fuente: Elaboración propia

Ratio respuesta	Cantidad de host
100	3640
99	169
98	65
97	49
96	90
95	64
94	63
93	72
92	36
91	46
90	181

Ratio respuesta	Cantidad de host
18	1
13	4
0	136

Realizando la suma de la cantidad de *hosts* que se sitúan dentro del rango del 90% al 100% de respuesta, se encuentran 4415 respecto a los 6025 que hay en total. Esto supone que hay 1610 *hosts* cuyo ratio de respuesta es inferior al 90%.

Por otro lado, se va a realizar una agrupación de todos los ratios de aceptación iguales y se calcula la media de todos los días que llevan en la plataforma los respectivos *hosts* (Tabla 12).

Tabla 12: Días medios del host en Airbnb agrupado por ratio de aceptación. Fuente: Elaboración propia

Ratio aceptación	Días medios en la plataforma
99	1858.117764
98	2122.985612
97	1903.730612
...	...
10	1629.000000
0	2204.485294

Al igual que en el caso anterior, la Figura 12 representa los valores para poder visualizarlos de forma gráfica.

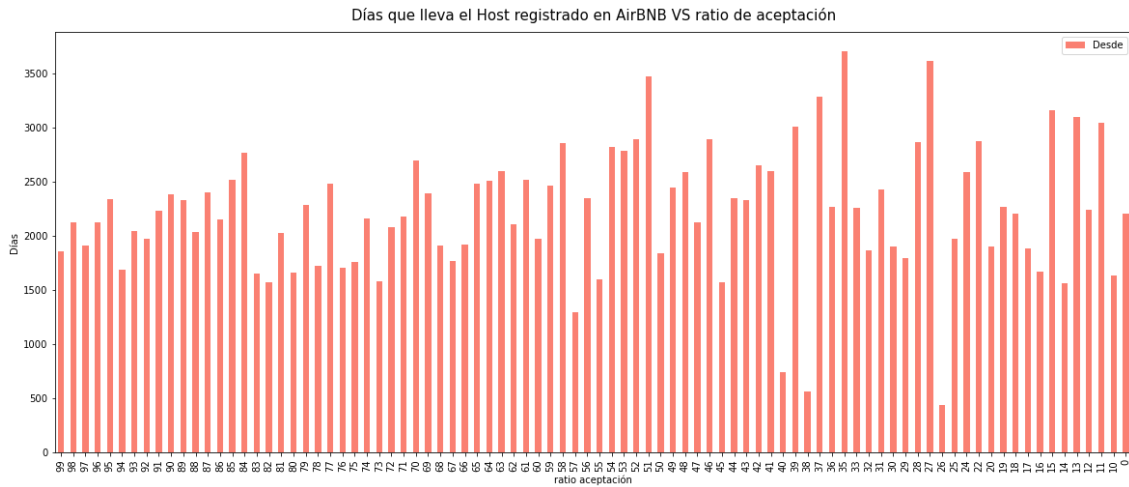


Figura 12. Gráfica de días medios que lleva el host en Airbnb por el ratio de aceptación (Fuente: Elaboración propia)

En este caso, sucede lo mismo que en la figura anterior, el grupo con la mayor cantidad de días en la plataforma y el grupo con menor días están relativamente cerca en cuanto a ella ratio de aceptación. Por lo que, no se puede asegurar una correlación alguna los días que lleve un *host* dado de alta en la plataforma con el ratio de aceptación.

Para finalizar con el análisis de los *hosts*, se procede a analizarla relación entre la cantidad de días que lleva el *host* en la plataforma y si es o no *superhost*. Para ello, se agrupan todos los valores *true* (que indican que sí que son *superhost*) y todos los valores *false* (que indica que no son *superhost*) y se calcula la media de cada uno de ellos.



Figura 13. Gráfica de días medios que lleva el host en Airbnb por su condición de *superhost* (Fuente: Elaboración propia)

Como se puede observar en la Figura 13, la media de los *superhost* es mayor a la de los que no son *superhost*, por lo que muestra que, por lo general, un *host* para llegar a ser *superhost* debe permanecer más tiempo en la plataforma.

En la siguientes tablas (Tablas 13, 14y 15) se continúa analizando la información relativa a los *superhost*. Se van a estudiar tres aspectos: (i) para alquileres de casas enteras; (ii) alquileres de habitaciones privadas, y (iii) alquileres de habitaciones compartidas; considerando el género del *host*, la cantidad de reseñas, la media de cada una de las reseñas y si es *superhost*.

El primer paso consiste en comprobar si existe relación alguna entre *superhost* y la cantidad y calidad de las reseñas y, por otro lado, la relación entre la cantidad de hombres y mujeres que son o no *superhost*.

La Tabla 13 relaciona los alquileres de casas enteras con los factores mencionados anteriormente.

Tabla 13: Cantidad de alquileres, reseñas valor medio de reseñas por género y por tipo de *superhost* de casas enteras. Fuente: Elaboración propia

¿ <i>Superhost</i> ?	Género	Cantidad de alquileres	Cantidad de reseñas totales	Valor medio de reseñas por casa	Cantidad media de reseñas por casa
f	Grupo	405	14423	4.421224	35.612346
	Hombre	1524	53420	4.516131	35.052493
	Mujer	1187	42038	4.574514	35.415333
t	Grupo	90	11778	4.827727	130.866667
	Hombre	549	46571	4.836660	84.828780
	Mujer	388	26567	4.844282	68.471649

Ante la pregunta de si había alguna relación entre ser o no *superhost* y la cantidad y calidad de las reseñas, se puede observar en la Tabla 13, que por un lado, el valor medio de las reseñas por casa es superior en los usuarios que son *superhost* de los que no lo son. Por otro lado, la cantidad media de reseñas por casa también es superior en los 3 casos.

Otra de las preguntas planteadas que se quiere analizar es si hay alguna relación entre la cantidad de hombres y mujeres que son o no *superhost*. En base a los resultados mostrados en la Tabla 13, un 36% de los hombres son *superhost* mientras que el 32% de las mujeres son *superhost*. Aunque hay diferencia entre ambos géneros no es significativa para afirmar que los hombres tienen una mayor cantidad de cuentas *superhost* de manera significativa.

La Tabla 14 relaciona los alquileres de habitaciones privadas con el género del *host*, la cantidad de reseñas, la media de cada una de las reseñas y si es *superhost*.

Tabla 14: Cantidad de alquileres, reseñas valor medio de reseñas por género y por tipo de *superhost* de habitaciones privadas. Fuente: Elaboración propia

¿Superhost?	Género	Cantidad de habitaciones privadas en alquiler	Cantidad de reseñas totales de las habitaciones privadas	Valor medio de reseñas por habitación privada	Cantidad media de reseñas por habitación privada
f	Grupo	66	1455	4.569130	22.045455
	Hombre	825	21201	4.472372	25.698182
	Mujer	600	10963	4.495054	18.271667
t	Grupo	18	721	4.756875	40.055556
	Hombre	172	7604	4.890921	44.209302
	Mujer	146	7668	4.852552	52.520548

En cuanto a la relación entre ser o no *superhost* y la cantidad y calidad de las reseñas, se puede observar a través de la Tabla 14, que el valor medio de las reseñas por casa es muy superior en los *hosts* que son *superhost* de los que no lo son. Por otro lado, la cantidad media de reseñas por casa también es superior en los 3 casos, llegando a duplicar en el género femenino.

Respecto a si hay alguna relación entre la cantidad de hombres y mujeres que son o no *superhost*, destacar que un 21% de los *superhost* son hombres mientras que el 24% son mujeres. Aunque hay diferencia entre ambos géneros no es significativa para afirmar que las mujeres tienen una mayor cantidad de cuentas *superhost* de manera significativa.

La Tabla 15 relaciona los alquileres de habitaciones compartidas con el género del *host*, la cantidad de reseñas, la media de cada una de las reseñas y si es *superhost*.

Tabla 15: Cantidad de alquileres, reseñas valor medio de reseñas por género y por tipo de superhost de habitaciones compartidas. Fuente: Elaboración propia

¿Superhost?	Género	Cantidad de habitaciones compartidas en alquiler	Cantidad de reseñas totales de las habitaciones compartidas	Valor medio de reseñas por habitación compartida	Cantidad media de reseñas por habitación compartida
f	Hombre	30	64	4.538462	2.133333
	Mujer	5	9	4.933333	1.800000
t	Mujer	1	0	0.000000	0.000000

En este caso, solo hay 1 host que es *superhost* y, por este motivo, no pueden realizarse comparaciones.

3.4.3 Género

En este apartado se analiza si existe alguna relación entre si el *host* es hombre, mujer o grupo y el hecho de que éste tenga una mayor o menor cantidad de licencias.

Para ello, se desarrollan 3 análisis (uno para alquileres de casas enteras, otro para alquileres de habitaciones privadas y el último para alquileres de habitaciones compartidas) que contenga de todas los alquileres que hay, su género y si tiene o no licencia (Tabla 16 reducida para la memoria).

En este sentido la Tabla 16 está relacionada con las licencias y géneros de alquileres de casas enteras.

Tabla 16: Género del host y licencia. Fuente: Elaboración propia

	Género	Licencia
0	Hombre	VT-41540-V
1	Mujer	NaN
2	Mujer	31815
3	Hombre	NaN
4	Hombre	VT-36438-V

Todo aquel *host* que tenga un *NaN* en su licencia significa que no tiene licencia.

Para poder estudiar la información de cada género, de forma clara, se agrupa por género y se realiza un conteo de la cantidad de licencias no nulas (no *NaN*) que tiene cada uno.

Tabla 17: Cantidad de host y licencias por género en casas enteras . Fuente: Elaboración propia

Género	Cantidad de Hosts	Licencias	% de Licencias respecto al total del género
Grupo	495	307	62.020202
Hombre	2073	833	40.183309
Mujer	1575	615	39.047619

La tabla 17 muestra la información referente a los alquileres de casas completas, en la cual se puede apreciar que, en primer lugar hay más hombres que tienen una mayor cantidad de propiedades enteras, y con ello de licencias, pero en cuanto al porcentaje de licencias respecto al total de ambos géneros, tanto hombres como mujeres, presentan un porcentaje similar (cerca al 40%). En cuanto a los grupos, es lógico que presenten una mayor cantidad de licencias debido a que normalmente tienen mayores recursos para gestionar las licencias.

La Tabla 18 está relacionada con las licencias y géneros de alquileres de habitaciones privadas.

Tabla 18: Cantidad de host y licencias por género en habitaciones privadas. Fuente: Elaboración propia

Género	Cantidad de Hosts	Licencia	% de Licencias respecto a la q de host
Grupo	84	14	16.666667
Hombre	997	46	4.613842
Mujer	746	38	5.093834

La Tabla 18 muestra la información referente a los alquileres de habitaciones privadas, y se puede apreciar que, al igual que en la anterior tabla, hay más hombres que tienen una mayor cantidad de propiedades enteras, y con ello de licencias, pero en cuanto al porcentaje de licencias respecto al total de ambos géneros, tanto hombres como mujeres presentan un porcentaje similar (cerca al 5%). Respecto a los grupos, es natural que presenten una mayor cantidad de licencias debido a que normalmente tienen mayores recursos para gestionar las licencias.

La Tabla 19 está relacionada con las licencias y géneros de alquileres de habitaciones compartidas.

Tabla 19: Cantidad de host y licencias por género en habitaciones compartidas. Fuente: Elaboración propia

Género	Cantidad de Hosts	Licencia	% de Licencias respecto a la q de host
Hombre	30	3	10.0
Mujer	6	0	0.0

La Tabla 19 muestra la información referente a los alquileres de habitaciones compartidas, y se puede apreciar que, al igual que en la anterior tabla hay más hombres que tienen una mayor cantidad de habitaciones compartidas, y con ello de licencias, ya que las mujeres en este tipo de propiedad no presentan licencias. Respecto a los grupos, se puede apreciar que no tienen ninguna propiedad de este tipo.

En las siguientes tablas se continúa analizándola información que el *dataset* ofrece en relación al género. En este caso, se analizan las relaciones que se presentan entre el género y el tipo de casas que tienen (casas enteras, habitaciones privadas y habitaciones compartidas).

Para ello, se estudiará el género del *host*, la cantidad de propiedades que tiene de cada tipo de *host* y la *url* por si se desea corroborar alguna propiedad.

Tabla 20: Nombre del *host* junto con su género y la cantidad de alquileres que tiene de cada tipo. Fuente: Elaboración propia

	Nombre del host	Género	Total de casas/hab en alquiler	casas enteras en alquiler	hab priv en alquiler	hab comp en alquiler	url
0	Toni	Hombre	2	2	0	0	https://www.airbnb.com/users/show/219476
1	Pilar	Mujer	1	1	0	0	https://www.airbnb.com/users/show/120649
2	Carmen	Mujer	5	5	0	0	https://www.airbnb.com/users/show/508042
3	Santiago	Hombre	1	1	0	0	https://www.airbnb.com/users/show/1451371
4	Vicente	Hombre	7	6	1	0	https://www.airbnb.com

Para poder analizar si existe algún tipo de correlación o anomalía, se agrupa la información de la Tabla 20 por género y se suman las distintas propiedades que tienen por tipo.

Tabla 21: Cantidad de alquileres que tiene de casa complicada *host* agrupado por género. Fuente: Elaboración propia

Género	Usuarios	Cantidad de casas enteras	Cantidad de habitaciones privadas	Cantidad de habitaciones compartidas	Total alquileres por Género
Grupo	171	495	84	0	579
Hombre	1726	2073	997	30	3100
Mujer	1442	1575	746	6	2327

Género	Usuarios	Cantidad Media de casas enteras	Cantidad Media de habitaciones privadas	Cantidad Media de habitaciones compartidas
Grupo	171	2.894737	0.491228	0.000000
Hombre	1726	1.201043	0.577636	0.017381
Mujer	1442	1.092233	0.517337	0.004161

Como se puede observar en la Tabla 20, en valores absolutos, hay una mayor cantidad de hombres que poseen propiedades en la plataforma, pero si se observan los valores porcentuales de la Tabla 21 se puede observar que prácticamente ambos géneros tienen las mismas casas.

En cuanto a los grupos, hay que destacar que los mismos presentan una mayor cantidad porcentual de casas enteras, pero de habitaciones presentan una menor cantidad, como cabría esperar.

la Figura 14 muestra los valores obtenidos en las tablas anteriores, representados gráficamente.

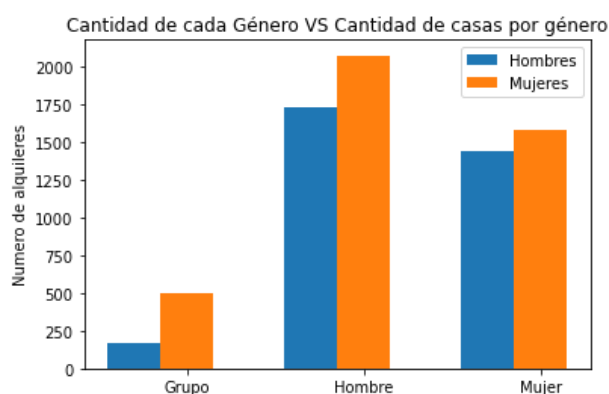


Figura 14. Gráfica de la relación entre géneros. (Fuente: Elaboración propia)

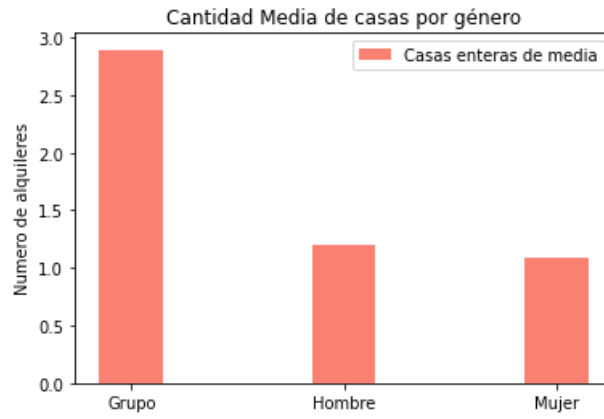


Figura 15. Gráfica de la cantidad media de casas por género (Fuente: Elaboración propia)

3.4.4 Barrios

Una de las grandes divisiones que se puede hacer es mediante la localización de los alojamientos por barrios. Como se ha comentado anteriormente, desde Airbnb ofrecen los datos tanto del barrio en el que está ubicado el alojamiento como el sub-barrio.

Estos son los sub-barrios con los que se va a trabajar en esta sección:

MORVEDRE, LA SEU, EL CARME, ARRANCAPINS, LA MALVA-ROSA, EL PILAR, EL MERCAT, MONT-OLIVET, CABANYAL-CANYAMELAR, EL CARME, MONT-OLIVET, LA ROQUETA, EL MERCAT, NOU MOLES, AIORA, EL CARME, CABANYAL-CANYAMELAR, CABANYAL-CANYAMELAR, MONT-OLIVET, LA CREU DEL GRAU, EXPOSICIO, RUSSAFA, LA MALVA-ROSA, LA CARRASCA, MORVEDRE, EL BOTANIC, RUSSAFA, LA SEU, LA MALVA-ROSA, EL CARME, EL CARME, EL CARME, EL PILAR, LA CREU DEL GRAU, ARRANCAPINS, CIUTAT DE LES ARTS I DE LES CIENCIES, RUSSAFA, SANT FRANCESC, SANT FRANCESC, EL MERCAT, SANT FRANCESC, EL MERCAT, ARRANCAPINS, LA MALVA-ROSA, BENICALAP, LA CARRASCA, PENYA-ROJA, SANT PAU, SANT FRANCESC, SANT FRANCESC, LA CARRASCA, AIORA, PATRAIX, ARRANCAPINS, CAMI FONDO, LA SEU, EXPOSICIO, EL PILAR, LA CARRASCA, CABANYAL-CANYAMELAR, TRINITAT, EL PILAR, EL MERCAT, EL PILAR, SANT PAU, LA SEU, EN CORTS, RAFALELL-VISTABELLA, PENYA-ROJA, RUSSAFA, EL PILAR, CIUTAT DE LES ARTS I DE LES CIENCIES, EL MERCAT, SANT ANTONI, MALILLA, MESTALLA, EL GRAU, EL PERELLONET, RUSSAFA, CABANYAL-CANYAMELAR, EL BOTANIC, EL MERCAT, MONT-OLIVET, RUSSAFA, RUSSAFA, EL SALER, EL MERCAT, EL MERCAT

Estos son los barrios con los que se va a trabajar en esta sección:

ALGIROS, BENICALAP , BENIMACLET, CAMINS AL GRAU, CAMPANAR, CIUTAT VELLA, EL PLA DEL REAL, EXTRAMURS, JESUS, L'EIXAMPLE, L'OLIVERETA, LA SAIDIA, PATRAIX, POBLATS DE L'OEST, POBLATS DEL NORD, POBLATS DEL SUD, POBLATS MARITIMS, QUATRE CARRERES, RASCANYA

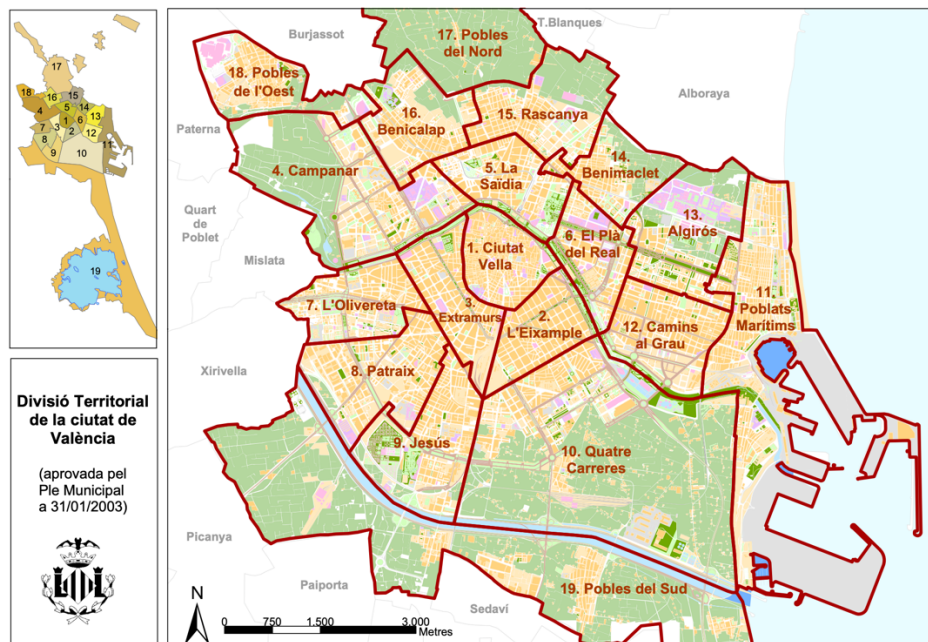


Figura 16. Barrios de Valencia (Fuente: www.valencia.es)

La Tabla 22 muestra la información compuesta por la agrupación de los barrios y la media de habitaciones, camas y precio que tiene cada uno.

Tabla 22: Cantidad de habitaciones, camas y precio medio agrupado por barrio. Fuente: Elaboración propia

Barrio Gr	Habitaciones	Camas	Precio
ALGIROS	2.254902	2.488095	65.958015
BENICALAP	1.425000	1.641304	52.382979
BENIMACLET	1.891304	2.222222	83.70212
...
POBLATS DE L'OEST	1.666667	2.700000	93.295082
POBLATS DEL NORD	2.015873	3.187500	131.109375
POBLATS DEL SUD	2.348101	3.500000	104.110429
POBLATS MARITIMS	1.948571	2.695612	93.170886

Presentando los resultados de forma tabular, tal y como se muestran en la Tabla 22, se hace difícil analizar si existe alguna relación entre las diferentes variables. Por ello, se procede a realizar una correlación entre las columnas de la Tabla 22.

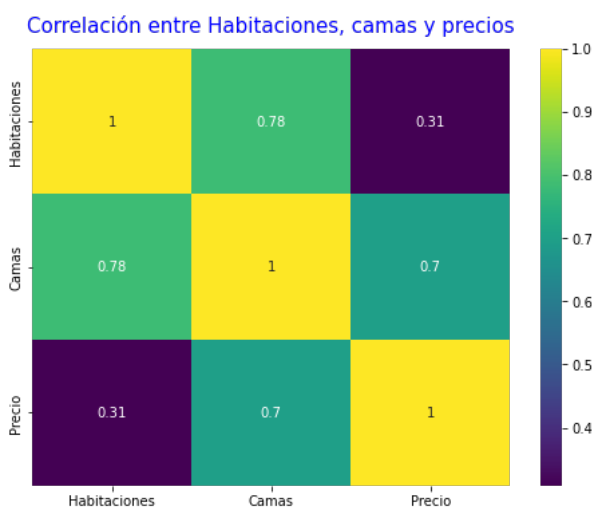


Figura 17. Correlación entre habitaciones, camas y el precio (Fuente: Elaboración propia)

Como se puede observar, existen valores entre el 0,7 y el 1 en las combinaciones de columnas:

- Camas y habitaciones
- Camas y precio

Esto indica que si aumenta el número de camas, hay una correlación positiva que muestra que también aumenta la cantidad de habitaciones y el precio.

La Tabla 23 muestra los barrios y las licencias. En este caso se pretende comprobar la cantidad de licencias que tiene cada barrio. Para ello, se realiza una agrupación de los barrios y se suma la cantidad de licencias que hay en cada uno de ellos.

Tabla 23: Cantidad de licencias agrupado por barrio. Fuente: Elaboración propia

Barrio Gr	Total de licencias
ALGIROS	37
BENICALAP	16
BENIMACLET	14
CAMINS AL GRAU	115
CAMPANAR	18
...	
CIUTAT VELLA	561
EL PLA DEL REAL	28
POBLATS MARITIMS	382
QUATRE CARRERES	112
RASCANYA	13

Para poder observar, de manera visual, la cantidad de licencias por barrio, el primer paso es ordenar los barrios mediante un 'sort()'. De esta manera queda ordenado de forma descendente tal y como muestra la Figura 18.

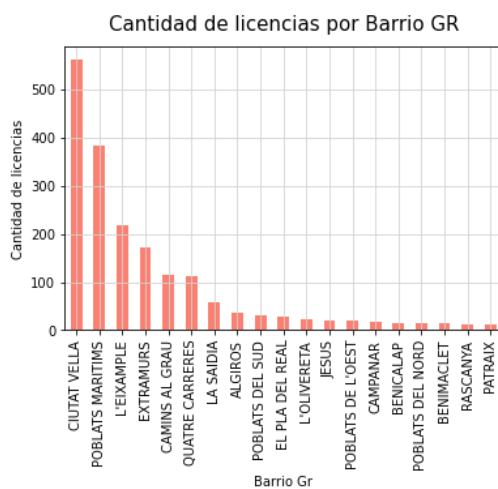


Figura 18. Gráfica de la cantidad de licencias que hay por barrio (Fuente: Elaboración propia)

De la Figura 18 se extrae que los barrios con un mayor volumen de licencias es Ciutat Vella, Poblat Marítims, L'eixample y Extramurs. Los 4 barrios están situados, bien en el centro de la ciudad, bien en la zona de la costa.

Para comprobar si existe alguna relación entre el número de licencias con la cantidad de alquileres que hay por barrio, se estudia la cantidad de alquileres que hay por barrio y se calcula el porcentaje que suponen las licencias respecto al total de alquileres que hay (Tabla 24 reducida para la memoria).

Tabla 24: Cantidad de alquileres agrupado por barrio. Fuente: Elaboración propia

Barrio Gr	Total Alquileres
ALGIROS	262
BENICALAP	94
BENIMACLET	94
CAMINS AL GRAU	451
CAMPANAR	84

Una vez obtenidos los valores totales de alquileres que hay por barrios, se deben combinar las tablas 23 y 24 para el cálculo del porcentaje (Tabla 25 reducida para la memoria).

Tabla 25: Cantidad de alquileres, licencias y el porcentaje de estas agrupado por barrio. Fuente: Elaboración propia

Barrio Gr	Total Alquileres	Total de licencias	Porcentaje
CIUTAT VELLA	1233	561	45.498783
L'EIXAMPLE	596	217	36.409396
POBLATS MARITIMS	1106	382	34.538879
EXTRAMURS	506	173	34.189723
POBLATS DE L'OEST	61	19	31.147541

Como se puede observar, Ciutat Vella no solo es el barrio dónde más alquileres hay disponibles, sino que también es el barrio con más licencias por apartamentos (un 45%).

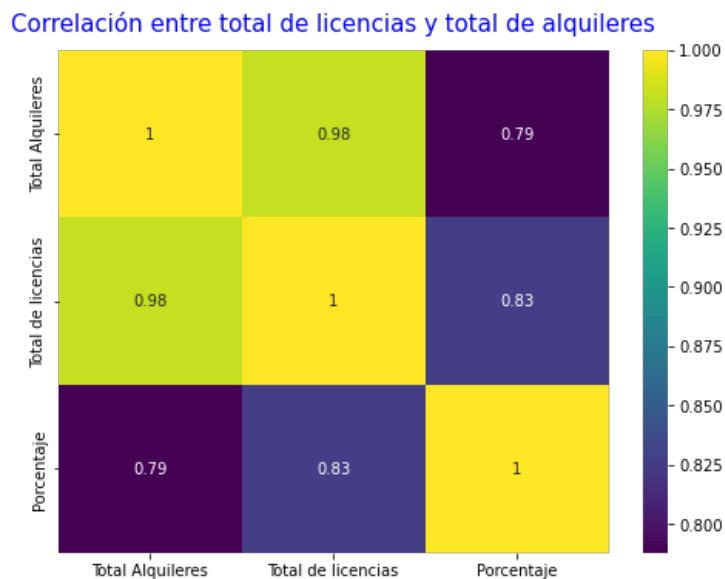


Figura 19. Correlación entre el total de licencias y el total de alquileres (Fuente: Elaboración propia)

La Figura 19 muestra que hay una correlación positiva entre el total de licencias y el total de alquileres que hay en un barrio. Aun embargo, a pesar de ser una correlación positiva tiene un valor porcentual bajo.

La Tabla 26 analiza si hay alguna relación entre los factores cama, disponibilidad y precios. Dicha tabla está compuesta por el barrio, las camas, la disponibilidad de los últimos 30, 60, 90 y 365 días y el precio.

Cabe destacar que la disponibilidad de los últimos 30, 60, 90 y 365 días significa que también tenemos la cantidad de veces que se ha reservado en esos periodos.

Tabla 26: Cantidad de camas, su disponibilidad y el precio. Fuente: Elaboración propia

	Barrio	Camas	Disponibilidad 30	Disponibilidad 60	Disponibilidad 90	Disponibilidad 365	Precio
0	MORVEDRE	2.0	4	21	51	326	66.0
1	LA SEU	1.0	0	0	0	263	100.0
2	EL CARMÉ	6.0	5	12	17	127	142.0
...
6023	LA CREU COBERTA	1.0	25	55	85	360	59.0
6024	MORVEDRE	1.0	22	52	82	95	25.0

En base a los resultados anteriores, se analiza, mediante un diagrama de caja y bigotes, las disponibilidades medias por barrios. Para ello, en primer lugar, se realiza la agrupación por barrios para su representación.

El factor de la media de días disponibles (30 y 60)

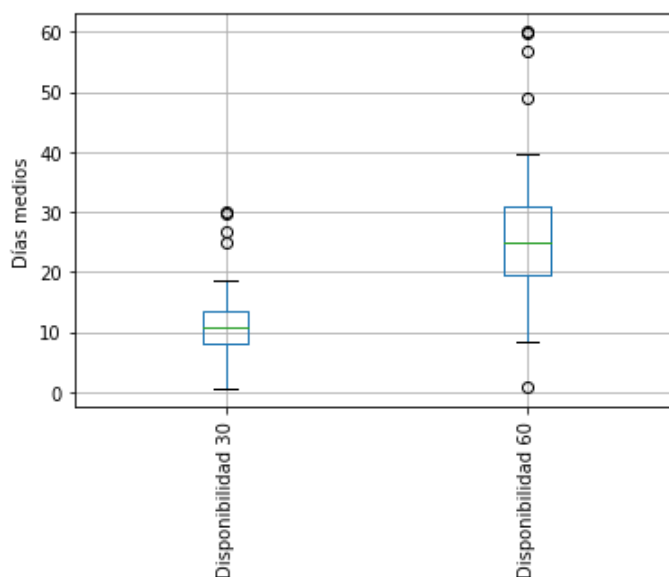


Figura 20. Diagrama de caja y bigotes de las disponibilidades de 30 y 60 días (Fuente: Elaboración propia)

La Figura 20 muestra que el grueso de las disponibilidades de los últimos 30 días ha rondado entre los 8 y los 13 días, y que ha presentado valores sospechosos por los 24 y los 30 días.

En cuanto a las disponibilidades de los 60 últimos días, las disponibilidades están concentradas entre los 19 y los 31, teniendo valores anormales en 1, 50 y 60.

Estos valores anormales pueden ser debidos a que puede haber apartamentos listados en varias plataformas y causar una serie de datos extraños.

El factor de la media de días disponibles (90 y 365)

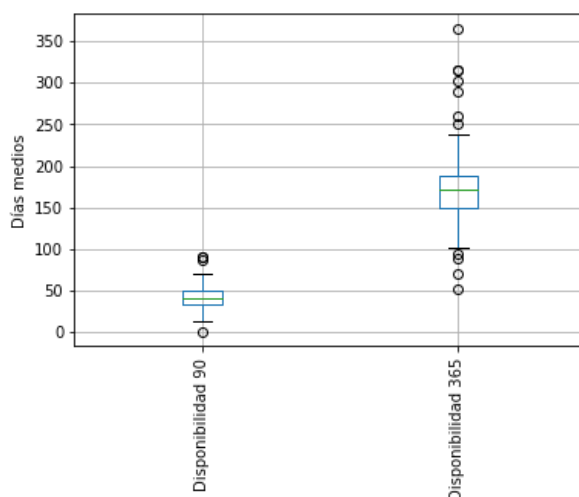


Figura 21. Diagrama de caja y bigotes de las disponibilidades de 90 y 365 días (Fuente: Elaboración propia)

En cuanto a las disponibilidades de los últimos 90 días se puede observar que el grueso de los días disponibles están concentrados entre los 35 y los 50 habiendo datos en los 90 y en el 0.

Finalmente, en la disponibilidad del último año, se observa que está concentrado entre los 150 y los 185 días. Por otro lado, ocurre una situación muy anómala y es la aparición de demasiados valores ilógicos. Esto puede ser debido a que hay muchas nuevas incorporaciones en Airbnb por lo que los apartamentos que estén menos de un año listados tendrán un número de días disponibles que no se corresponden a la realidad.

A continuación, las Figuras 22- 25 muestran un diagrama de caja y bigotes por cada una de las diferentes disponibilidades, pero en este caso, agrupada por barrios.

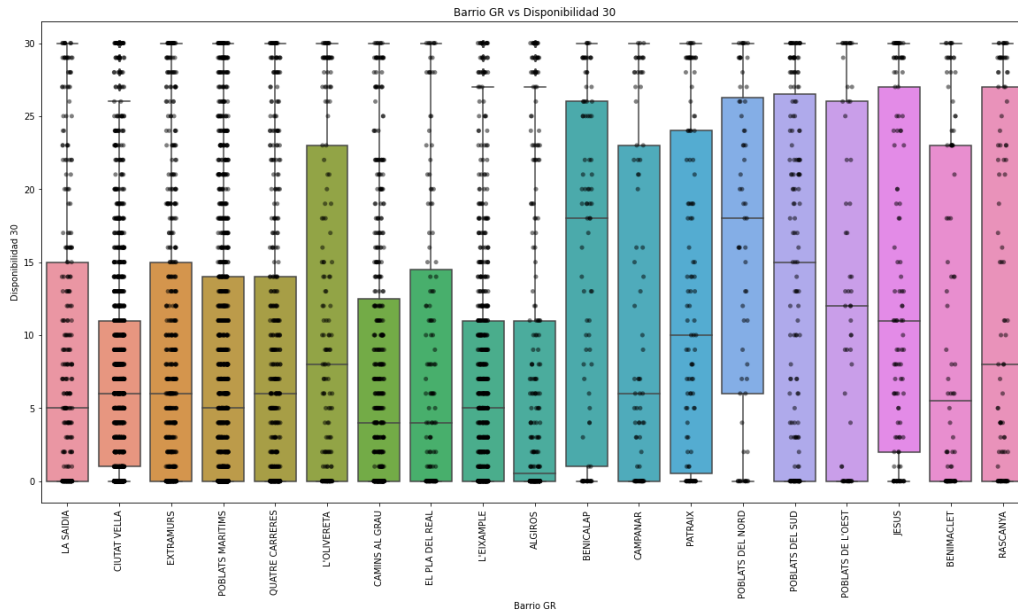


Figura 22. Diagrama de caja y bigotes de la disponibilidad de 30 días por barrios (Fuente: Elaboración propia)

Como se puede observar, los barrios con las disponibilidades más bajas en los últimos 30 días son Algirós y L'eixample, situándose entre los 0 y los 11 días disponibles. En cambio, los barrios con una disponibilidad media mayor son Poblats del Nord y Jesús.

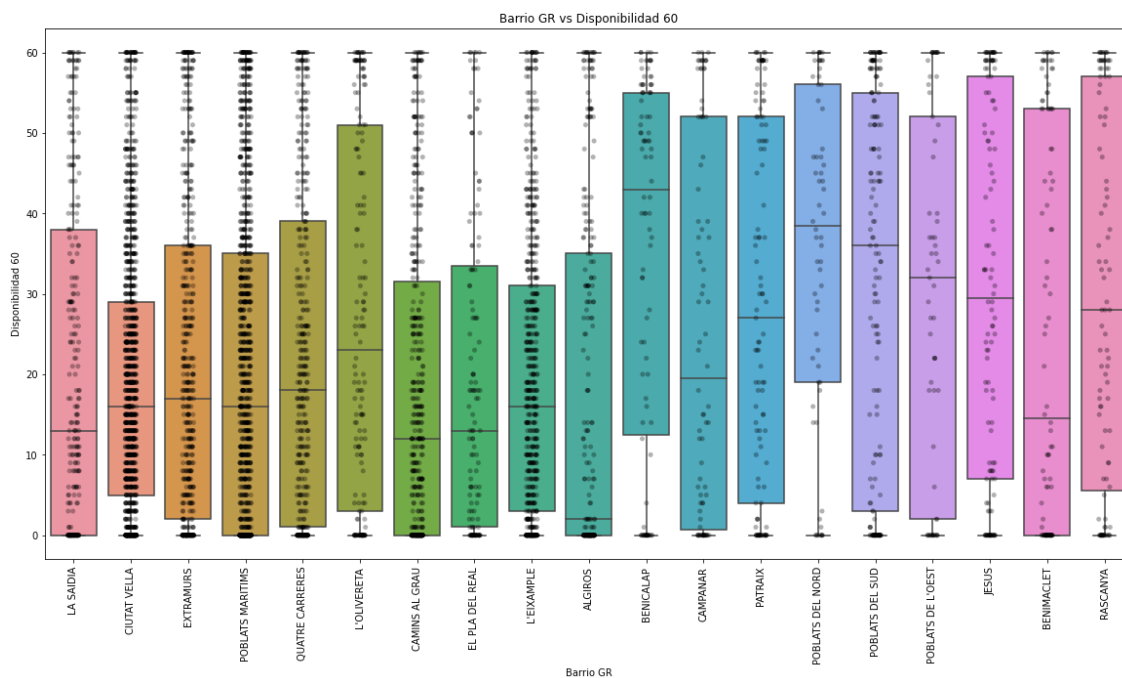


Figura 23. Diagrama de caja y bigotes de la disponibilidad de 60 días por barrios (Fuente: Elaboración propia)

En la representación de los últimos 60 días (Figura 23), los barrios con menor disponibilidad son camins al grau y L'eixample, situándose entre los 0 y los 32 días disponibles. Por otro lado, los barrios con mayor disponibilidad son Rascanya y Jesús.

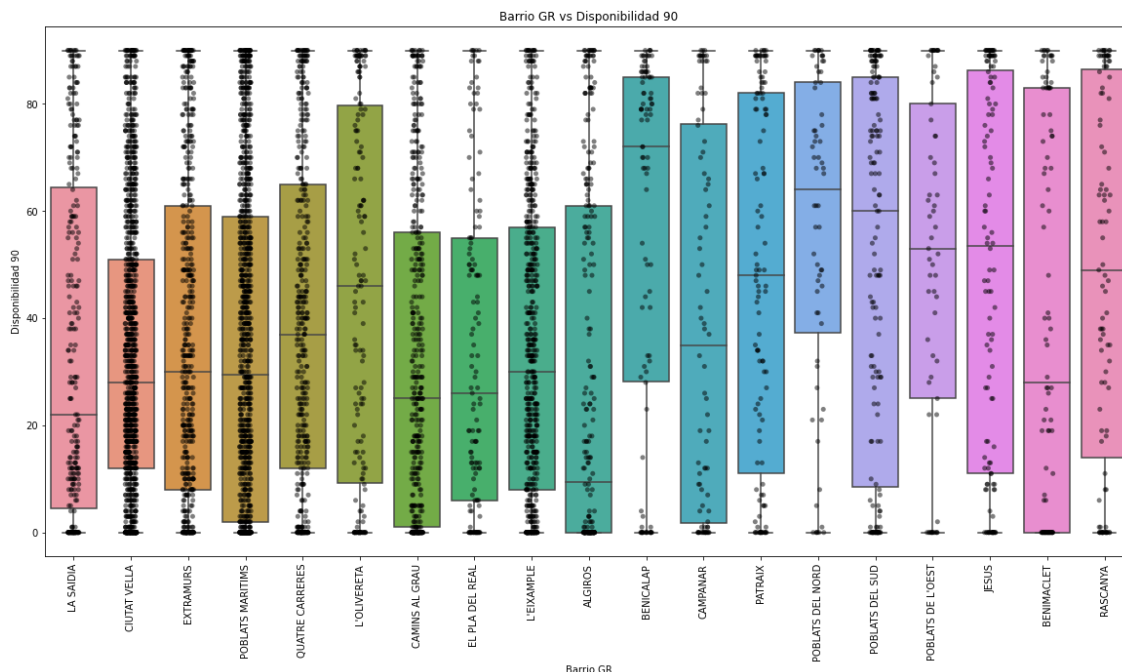


Figura 24. Diagrama de caja y bigotes de la disponibilidad de 90 días por barrios (Fuente: Elaboración propia)

Como se puede observar en la Figura 24, los barrios con las disponibilidades más bajas en los últimos 90 días son camins al grau y el pla del real, situándose entre los 0 y los 60 días

disponibles. En cambio, los barrios con una disponibilidad media mayor son Poblats del Nord y Benicalap.

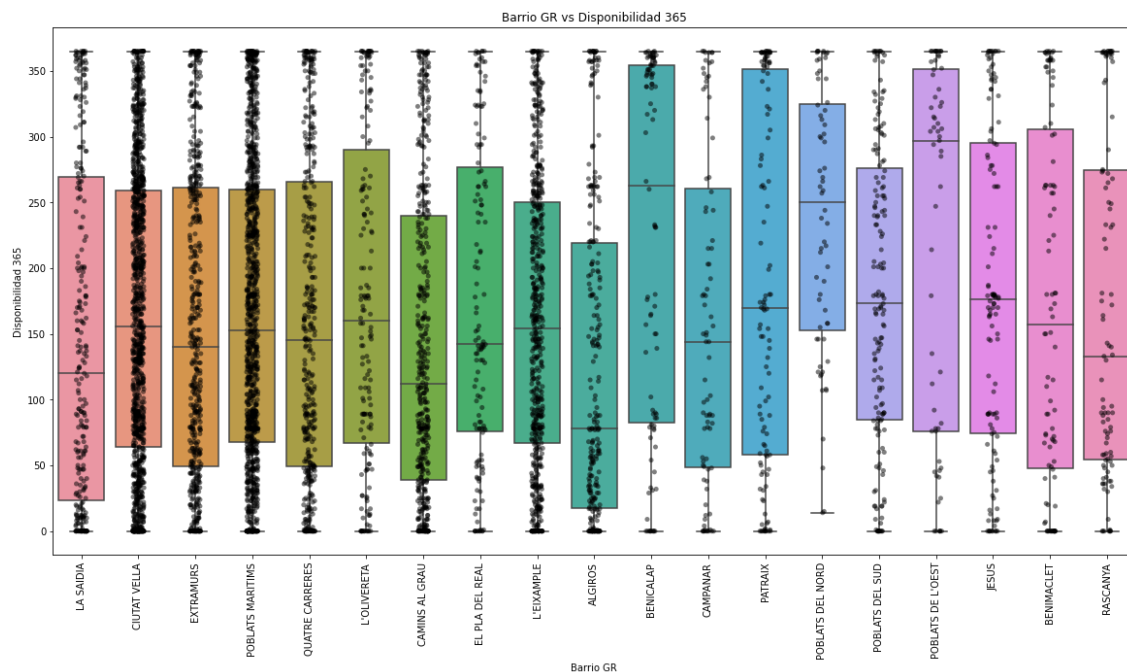


Figura 25. Diagrama de caja y bigotes de la disponibilidad de 365 días por barrios (Fuente: Elaboración propia)

En la representación del último año (Figura 25), los barrios con menor disponibilidad son Camins al Grau y Algirós, situándose entre los 25 y los 220 días disponibles. Por otro lado, los barrios con mayor disponibilidad son Benicalap y Poblats del Oest.

Para ver la progresión que ha sufrido la disponibilidad a lo largo del tiempo, se calcula la media por cada barrio.

Tabla 27: Disponibilidades absolutas y porcentuales agrupadas por barrio. Fuente: Elaboración propia

Barrio GR	Disponibilidad 30 mean	Disponibilidad 60 mean	Disponibilidad 90 mean	Disponibilidad 365 mean	%30 mean	%60 mean	%90 mean	%365 mean
ALGIROS	7.1297 71	17.351145	28.908397	124.40076 3	23.765 903	28.918 575	32.120 441	34.082 401
BENICALAP	14.968 085	34.648936	55.457447	220.39361 7	49.893 617	57.748 227	61.619 385	60.381 813
BENIMACLET	10.819 149	24.914894	39.670213	171.13829 8	36.063 830	41.524 823	44.078 014	46.887 205
CAMINS AL GRAU	7.6718 40	18.345898	31.640798	142.20399 1	25.572 801	30.576 497	35.156 442	38.959 998

Barrio GR	Disponibilidad 30 mean	Disponibilidad 60 mean	Disponibilidad 90 mean	Disponibilidad 365 mean	%30 mean	%60 mean	%90 mean	%365 mean
CAMPANAR	11.226 190	25.404762	39.273810	157.77381 0	37.420 635	42.341 270	43.637 566	43.225 701
CIUTAT VELLA	7.7785 89	18.971614	32.562855	162.47364 2	25.928 629	31.619 357	36.180 950	44.513 326
POBLATS DEL NORD	16.140 625	35.078125	55.937500	236.64062 5	53.802 083	58.463 542	62.152 778	64.833 048
POBLATS DEL SUD	14.104 294	31.417178	50.110429	178.74233 1	47.014 315	52.361 963	55.678 255	48.970 502
POBLATS MARITIMS	8.3164 56	20.375226	34.447559	164.60217 0	27.721 519	33.958 710	38.275 065	45.096 485
QUATRE CARRERES	8.9877 75	22.090465	38.643032	158.74327 6	29.959 250	36.817 441	42.936 702	43.491 309
RASCANYA	12.767 677	29.616162	48.020202	168.26262 6	42.558 923	49.360 269	53.355 780	46.099 350

La Tabla 27 muestra que no solo está la media de cada uno de los 4 periodos por cada barrio, sino que también está el porcentaje que representa respecto a su periodo, es decir, si son 15 días y está en el periodo de los 30 días será un 50%.

A simple vista, mediante las columnas %30 y %365 se puede observar una pequeña tendencia decreciente, es decir, que si se compara la tendencia que ha habido desde hace un año hasta ahora y la tendencia de este último mes, se aprecia que ahora hay menos días disponibles. Esto es algo positivo porque que haya menos días disponibles significa que hay una cantidad mayor de reservas.

Para ver la tendencia de forma gráfica se representa y se muestra en la Figura 26.

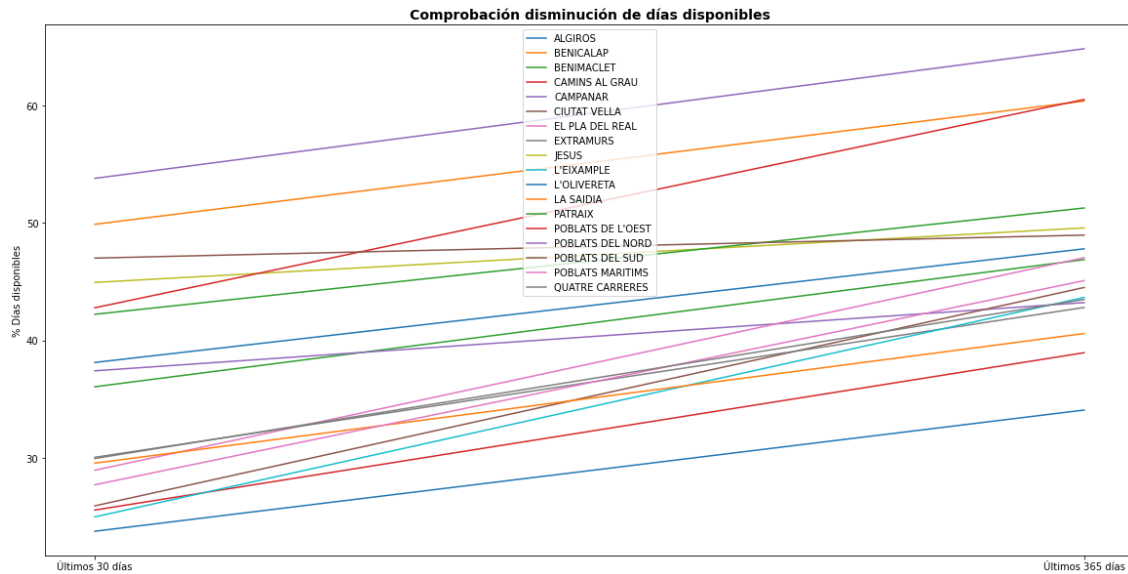


Figura 26. Gráfica de la tendencia de disponibilidad en el último año (Fuente: Elaboración propia)

Como se puede apreciar en la Figura 26, todos los barrios han tenido una disminución de días disponibles.

Por otro lado, para observar de manera precisa y visual las diferentes disponibilidades que tiene cada barrio se procede a representarlo gráficamente. La Figura 27 muestra la disponibilidad de 30 días por barrios; la Figura 28, la de 60 días; la Figura 29, la de 90 días; y finalmente, la Figura 30, la de todo el año.

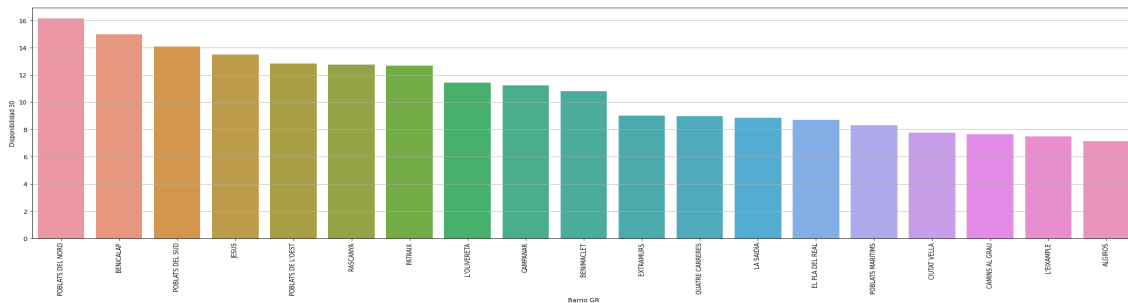


Figura 27. Gráfico de la disponibilidad de 30 días por barrios (Fuente: Elaboración propia)

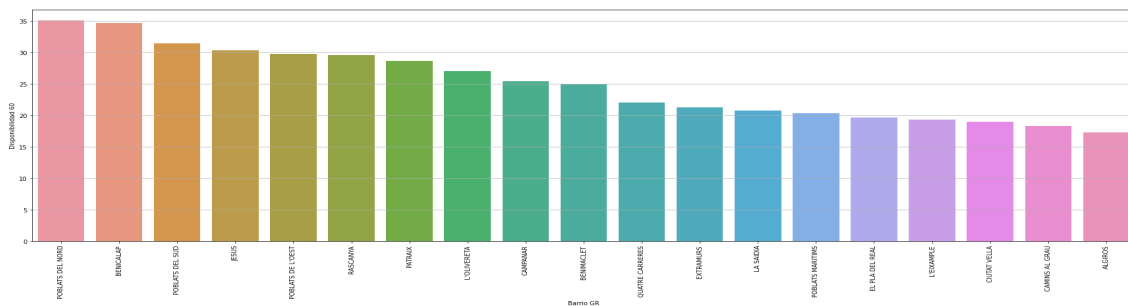


Figura 28. Gráfico de la disponibilidad de 60 días por barrios (Fuente: Elaboración propia)

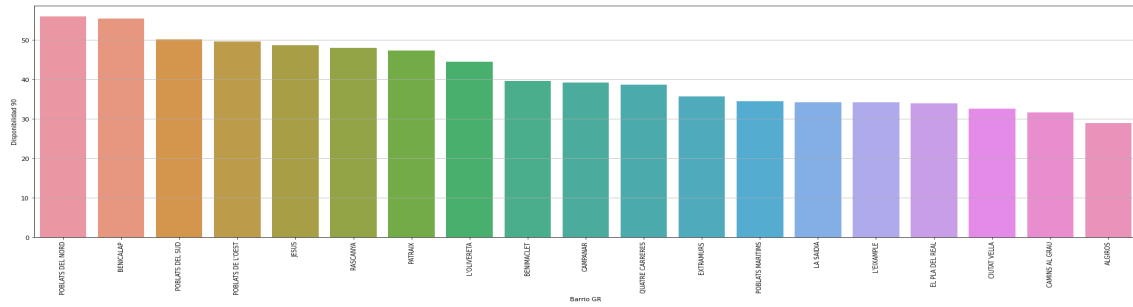


Figura 29. Gráfico de la disponibilidad de 90 días por barrios (Fuente: Elaboración propia)

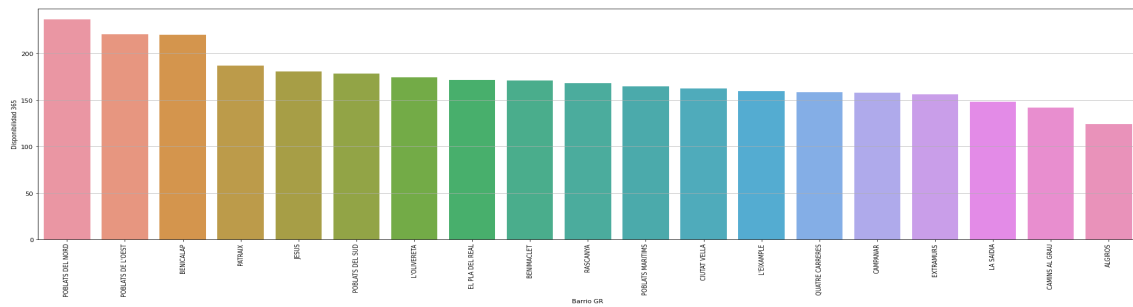
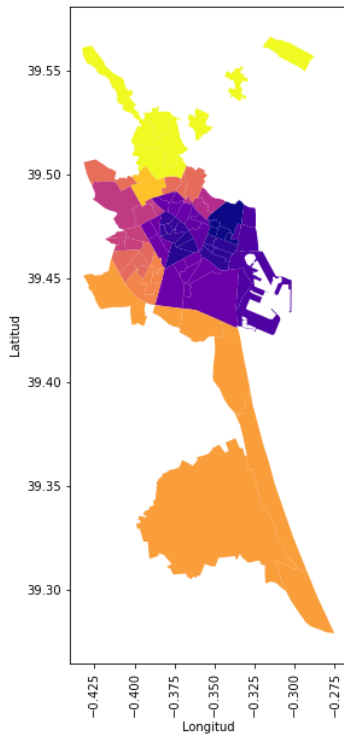


Figura 30. Gráfico de la disponibilidad de 365 días por barrios (Fuente: Elaboración propia)

Para facilitar el reconocimiento del barrio, se añade la geoposición, de forma que se pueda representar la disponibilidad en el mapa (Figuras 31 y 32).

Disponibilidad 30 media



Disponibilidad 60 media

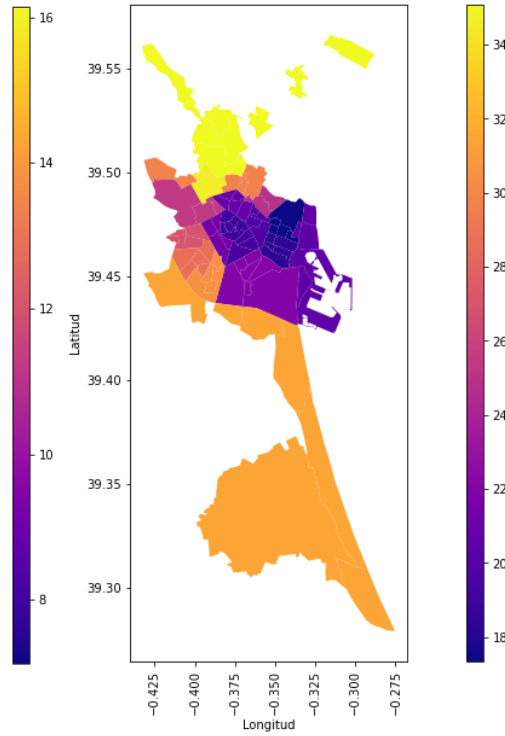


Figura 31. Representación en mapa de las disponibilidades medias de 30 y 60 días por barrios (Fuente: Elaboración propia)

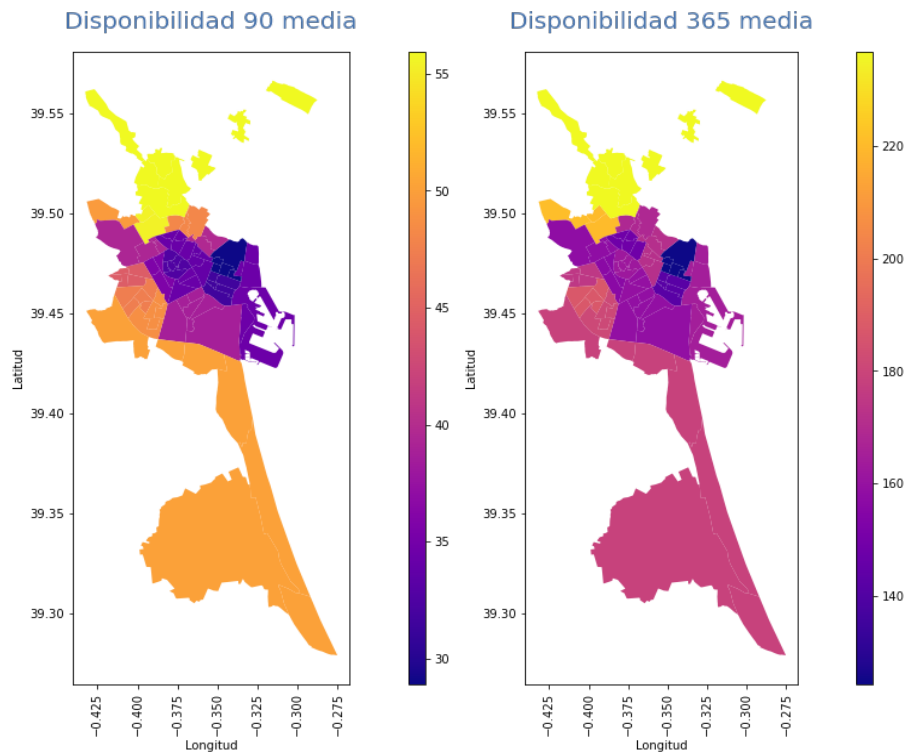


Figura 32. Representación en mapa de las disponibilidades medias de 90 y 365 días por barrios (Fuente: Elaboración propia)

Tal y como se puede observar, las zonas con menor disponibilidad son las zonas del centro de la ciudad y las de la costa.

A continuación, se analiza si existe alguna correlación entre el número de camas que hay en la vivienda con el precio que se cobra por ella.

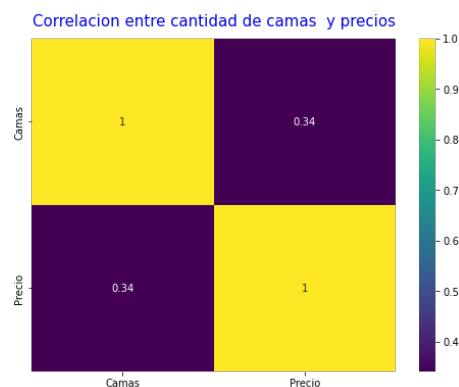


Figura 33. Correlación entre camas y precios (Fuente: Elaboración propia)

Como se puede observar en la Figura 33, no existe relación entre camas y precio.

Por otro lado, también se analiza si hay alguna correlación entre los días disponibles y el precio. Esto podría ser obvio en los billetes de avión (a más demanda, más caros; y a menos demanda,

más baratos) sin embargo, se tiene que analizar qué sucede en el caso de las reservas de propiedades.

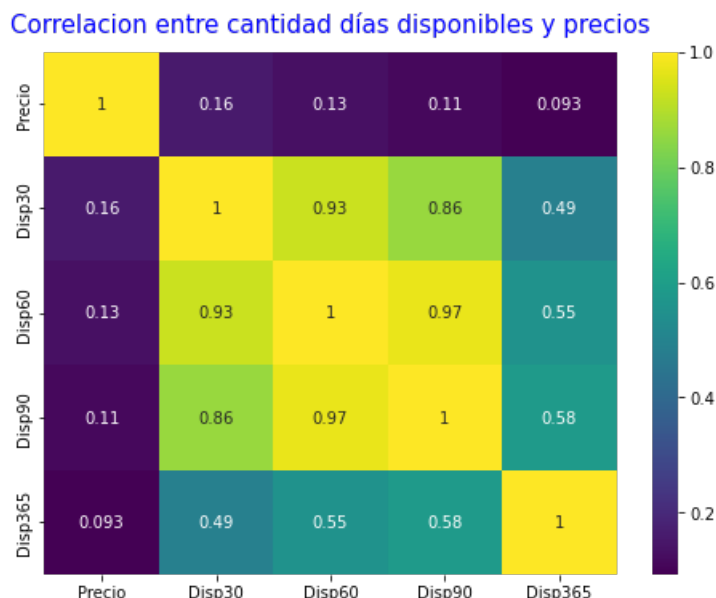


Figura 34. Representación de la correlación entre el precio y las distintas disponibilidades (Fuente: Elaboración propia)

En este caso, si se observa columna o fila precio, ya que en este momento no se desea analizar las correlaciones entre distintas disponibilidades, sino entre las disponibilidades y el precio; se aprecia que no existe relación alguna entre el precio y las distintas disponibilidades.

Finalmente, se analiza si existe alguna correlación entre la cantidad de camas y las disponibilidades, para corroborar si, a más camas, más disponibilidad (indicaría que los usuarios que reservan suelen ir solos o con poca compañía) o si a más camas, menos disponibilidad (indicaría que los usuarios que reservan suelen ir con mucha compañía).

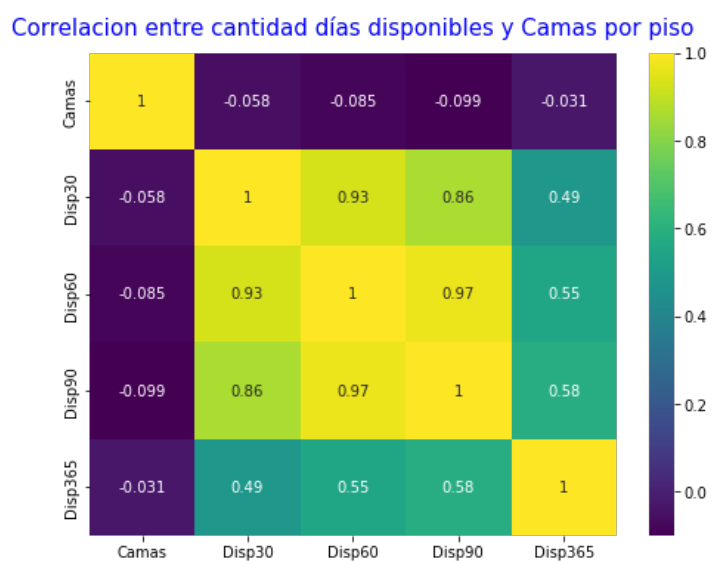


Figura 35. Representación de la correlación entre las camas y las distintas disponibilidades (Fuente: Elaboración propia)

Como se puede observar en la Figura 35, no hay una clara correlación entre el número de camas y la disponibilidad.

Por otro lado, cabría analizar en las casas con 1 habitación, la cantidad de camas que existen en dicha habitación. Esto se debe a que hay veces que los *hosts*, con tal de sacar un mayor rendimiento, aumentan el número de camas.

Tabla 28: Cantidad de camas en alojamientos con 1 habitación. Elaboración propia

Número de camas en 1 hab	Cantidad de camas
1.0	1935
2.0	772
3.0	115
4.0	30
5.0	4
6.0	5
7.0	2
8.0	2
9.0	1

Como se puede observar en la Tabla 28, hay una gran cantidad de habitaciones que tienen entre 1 y 4 camas. A pesar de ello, hay casas que tienen 1 habitación con más de 5 camas.

El mismo análisis, de la cantidad de camas, se realiza en las casas que poseen 2 habitaciones.

Tabla 29: Cantidad de camas en alojamientos con 2 habitaciones. Fuente: Elaboración propia

Número de camas en 2 hab	Cantidad de camas
1.0	29
2.0	552
3.0	611
4.0	212

Número de camas en 2 hab	Cantidad de camas
5.0	36
6.0	12
7.0	5
8.0	1
22.0	1

Como se puede observar en la Tabla 29, el grueso de habitaciones están agrupados en un rango correcto, aunque hay ciertas viviendas que presentan una cantidad de habitaciones muy altas.

Como se ha observado en el análisis, se han podido detectar 3 grandes grupos de barrios:

- Barrios de la zona de playa: Se trata de los barrios de Poblats Marítims, Camins al Grau y Algirós.
- Barrios del centro histórico: Se trata de los barrios de Ciutat Vella, Eixample, Extramurs, La saidia y El pla del Reial.
- Resto de barrios.

En el siguiente análisis se ha incluido otra columna, esta vez, de manera manual, a cuál de las 3 zonas pertenece el barrio. Junto con las zonas, también se han considerado las disponibilidades más extremas (últimos 30 y 365 días), las reseñas generales y el precio por cama.

Tabla 30: Reseñas, disponibilidades y zona del barrio agrupado por barrio. Fuente: Elaboración propia

Barrio GR	Reseña general	Disponibilidad últimos 30 d	Disponibilidad últimos 365 d	PrecioXCama	ZonaBarrio
ALGIROS	4.466000	5.707317	100.443902	26.072335	Barrios_Playa
BENICALAP	4.488429	13.814286	211.557143	36.164461	Barrios_Resto
BENIMACLET	4.553014	8.643836	154.986301	28.241784	Barrios_Resto
CAMINS AL GRAU	4.591899	6.820253	135.253165	34.618843	Barrios_Playa
CAMPANAR	4.631739	8.942029	146.304348	45.736567	Barrios_Resto
CIUTAT VELLA	4.578421	7.085873	155.228994	52.252199	Barrios_Centro
EL PLA DEL REAL	4.558522	7.113043	159.347826	29.892981	Barrios_Centro
EXTRAMURS	4.622558	8.088372	152.809302	45.634754	Barrios_Centro



Barrio GR	Reseña general	Disponibilidad últimos 30 d	Disponibilidad últimos 365 d	PrecioXCama	ZonaBarrio
JESUS	4.577444	11.622222	174.277778	37.636973	Barrios_Resto
L'EIXAMPLE	4.634747	7.197980	153.761616	47.040758	Barrios_Centro
L'OLIVERETA	4.565324	10.179856	168.848921	30.706107	Barrios_Resto
LA SAIDIA	4.590614	8.135965	144.719298	32.852007	Barrios_Centro
PATRAIX	4.560408	12.887755	187.153061	37.113684	Barrios_Resto
POBLATS DE L'OEST	4.647200	11.500000	207.760000	30.879723	Barrios_Resto
POBLATS DEL NORD	4.534318	16.568182	229.409091	25.145400	Barrios_Resto
POBLATS DEL SUD	4.530077	12.807692	172.500000	27.778385	Barrios_Resto
POBLATS MARITIMS	4.574565	7.733546	157.187898	40.154485	Barrios_Playa
QUATRE CARRERES	4.591729	8.665706	152.515850	37.917760	Barrios_Resto
RASCANYA	4.498481	12.101266	167.291139	44.253247	Barrios_Resto

Tras clasificar cada barrio, es importante realizar una agrupación por zonas, ya que de esta forma, se podrán observar las diferencias entre ellas.

Tabla 31: Cantidad media de reseñas, disponibilidades y precio medio por cama agrupado por zona. Fuente: Elaboración propia

ZonaBarrio	Media (1-5) Reseña General	Disponibilidad media últimos 30d	Disponibilidad media últimos 365d	Media PrecioXCama
Barrios_Centro	4.596972	7.524247	153.173407	41.534540
Barrios_Playa	4.544154	6.753705	130.961655	33.615221
Barrios_Resto	4.561651	11.612075	179.327603	34.688554

Respecto a la segunda columna de la Tabla 31, que se refiere al valor medio de reseñas que los usuarios ponen cuando reservan en cada zona, se puede apreciar que no hay un cambio significativo ya que la variación entre el máximo y el mínimo es de 0,05.

Por otro lado, en cuanto a las columnas de disponibilidades se observa una clara diferencia, siendo las dos zonas del centro y de la playa las que mayor reservas tienen. De estas dos, la que presenta una demanda más elevada es la de la playa.

Finalmente, el precio medio por cama por día es más elevado en el centro de la ciudad, dejando a las otras dos zonas con un precio muy similar.

Si se analiza el ámbito económico, se puede observar dónde se obtienen mayores ingresos de media por cama. Como se ha comentado antes, si hay disponibilidad, también hay reservas. Tan sólo resta calcular la diferencia entre el valor total (30 ó 365) y el valor de disponibilidad.

Tabla 32: Ingresos medios agrupados por zona. Fuente: Elaboración propia

ZonaBarrio	Ocupación media últimos 30d	Ocupación media últimos 365d	Media PrecioXCama(€)	Ingresos medios por cama 365 d (€)
Barrios_Centro	22.475753	211.8265	41.534540	8798.10776425
Barrios_Playa	23.246295	234.03835	33.615221	7867.20249675
Barrios_Resto	18.387925	185.672397	34.688554	6418,05863825

Tal y como se muestra en la Tabla 32, si el precio del inmueble es similar, tiene un mayor interés alquilar un piso en el centro más que en el resto de las zonas.

3.4.5 Reseñas

Para acceder a las reseñas, se precisa la obtención de los datos de otra tabla, en este caso, la llamada reviews.csv.gz. Esta tabla proporciona la información del *id* del alquiler, la fecha de la reseña, el *id* del usuario que ha hecho la reseña, y la reseña.

Tabla 33: Id del host, fecha, nombre del usuario y reseña. Fuente: Elaboración propia

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	48154	117554	2010-10-12	180238	Martha	Toni's place was perfect in so many ways. It ...
1	48154	145645	2010-11-28	204240	Mark	Awesome stay!! We'd recommend Toni's apartment...
2	48154	190572	2011-03-01	258565	Domenico	really nice house in a wonderful position! yo...
3	48154	195081	2011-03-08	213496	Romina & Martín	Apartamento muy agradable, al igual que su pro...
4	48154	218435	2011-04-05	340330	Jenna	Was a great apartment, easy access to the site...

Uno de los aspectos interesantes que ofrecen las reseñas, es conocer de dónde provienen los huéspedes, o en su defecto, qué idioma hablan. Para poder extraer dicha información, se debe aplicar a la columna de 'comments', el siguiente código:

- for i in df.index:
 - df.at[i,'comments']=langid.classify(df['comments'])[i][0]
 - fechapr=df['date'][i][0:4]
 - fechapr=int(fechapr)
 - df.at[i,'año']=fechapr

De esta manera, se obtiene el código del idioma que ha detectado. Estos son unos de los más usados (*Apéndice B. Idiomas — Lista de códigos ISO 639, s. f.*):

Tabla 34: Traducción nombres de idioma a ISO-639-1. Fuente: Elaboración propia

Nombre del idioma	ISO-639-1
Inglés	En
Español-Castellano	Es
Italiano	It
Francés	Fr
Alemán	De

Una vez conocidas algunas las abreviaturas, se procede a mostrar los resultados, en la tabla 35, de los idiomas.

Tabla 35: Id piso, id host, idioma y año de la reseña. Fuente: Elaboración propia

	listing_id	id	reviewer_id	comments	año
0	48154	117554	180238	en	2010.0
1	48154	145645	204240	en	2010.0
2	48154	190572	258565	en	2011.0
3	48154	195081	213496	es	2011.0
4	48154	218435	340330	en	2011.0
...
244807	581189171914728794	586051348230877101	430397409	fr	2022.0

	listing_id	id	reviewer_id	comments	año
244808	581337935125231192	585296203289281249	5765250	en	2022.0
244809	582406191849717400	584560436158338819	417355311	es	2022.0
244810	582433786124982535	586002624286044878	228263746	es	2022.0
244811	582525294233236075	586790189200905171	55866366	es	2022.0

Para poder analizar cuáles son los idiomas que más utilizan los usuarios, se realiza una agrupación por idioma y se realiza el conteo de la cantidad de reseñas que tiene (Tabla 36 reducida para la memoria)

Tabla 36: Cantidad de comentarios por idioma y su porcentaje agrupado por idioma. Fuente: Elaboración propia

Idioma	Cantidad de comentarios por idioma	Respecto al total(%)
en	119340	48.747610
es	67969	27.763753
fr	18094	7.390978
it	11711	4.783671
nl	8846	3.613385
de	8534	3.485940
ru	3108	1.269546
pt	2285	0.933369
zh	741	0.302681
pl	586	0.239367

En este caso, los idiomas más utilizados son el inglés, que representa casi el 50%; el español, con un poco más de ¼ de las reseñas; seguido por el francés, italiano, holandés, alemán, ruso, portugués, chino y polaco.

Cabe recalcar que estas son todas las reseñas que ha habido en Valencia desde que existe Airbnb. Para estudiar el progreso de las reseñas es necesario agruparlas por año.

Tabla 37: Cantidad de reseñas agrupado por año. Fuente: Elaboración propia

año	comments
2010	2

año	comments
2011	33
2012	295
2013	1212
2014	2714
2015	6402
2016	14366
2017	27136
2018	44401
2019	61684
2020	22772
2021	50496
2022	13299

En este caso para analizar los datos se va a eliminar el año 2022, debido a que no ha acabado y el año 2010 debido a que no empezó el día 1 de enero y no podrá ser comparado de la misma manera.

Para ver la evolución, se representa gráficamente (Figura 36 – 39), los diferentes idiomas utilizados los últimos 4 años.

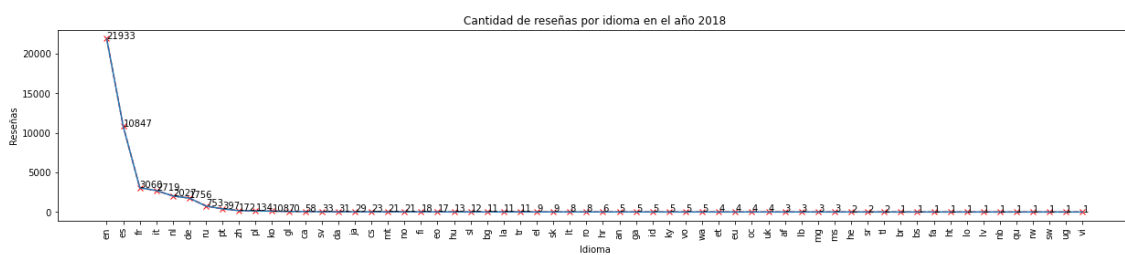
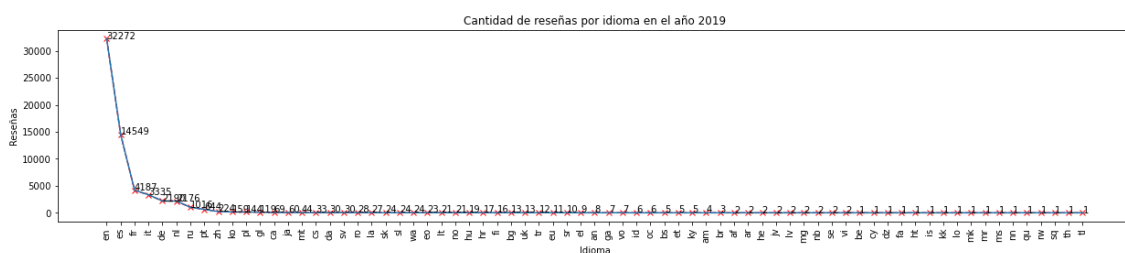


Figura 36. Representación gráfica de las reseñas de cada idioma en el año 2018 (Fuente: Elaboración propia)



Tal y como se puede observar en la Figura 40, no hay correlación alguna entre la cantidad de reseñas que se ponen por apartamento y el precio que se paga.

Para terminar el análisis, se realiza un estudio de la correlación del valor medio de las reseñas y el precio. Si antes se pretendía correlacionar el número de reseñas en función de su precio, ahora se busca conocer si los usuarios que pagan más, son más estrictos y ponen peores reseñas, o si por el contrario, los usuarios que pagan más, al alojarse en sitios supuestamente mejores, pone mejores reseñas.

Correlación entre el precio que se paga por reserva y el valor medio de las reseñas

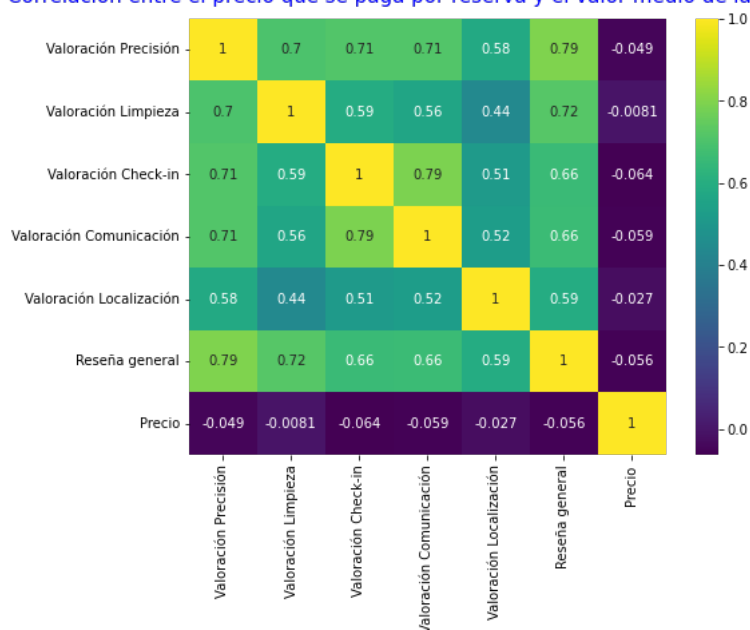


Figura 41. Representación de la correlación entre el precio y la valoración media de reseñas (Fuente: Elaboración propia)

En este caso, la última columna y/o fila, que es la que correlaciona el precio con el resto de las reseñas medias, muestra que tampoco hay correlación alguna.

3.5. Detección y propuesta de mejoras

Tras el análisis del caso de estudio, se han detectado una serie de puntos en los cuales se podrían implementar algunas mejoras, dividido según la clasificación del análisis realizado en el apartado 3.4.

3.5.1 En los tipos de alquiler

En el apartado de la representación del mapa de la cantidad de alquileres que hay por cada barrio, se ha podido detectar que gran parte de los pisos o habitaciones en alquiler están agrupadas en unas zonas muy concretas. Viendo que hay una parte de la ciudad que está tan masificada, convendría enfocar los esfuerzos en aconsejar a los *hosts* que sean más cuidadosos con la comunidad (tanto de vecinos como en el barrio), ya que en muchas ocasiones un gran número de personas distintas entrando y saliendo puede ser molesto para el resto de los propietarios. Esto se podría mejorar creando guías o minicursos de parte de Airbnb para ayudar a saber gestionar un piso de alquileres, mostrando cuántas horas deben dedicar a cada aspecto

de la gestión del alojamiento, las necesidades básicas que tiene que tener una casa en alquiler para no molestar a la comunidad, es decir, intentar profesionalizar este sector.

Por otro lado, en el número medio de reseñas que tiene cada apartamento, habitación u hotel son muy bajas. Las casas enteras, que son las que presentan un mayor número medio de reseñas, poseen, de media, 47 reseñas; mientras que el resto de los tipos de alquiler poseen de media 27 reseñas por habitación privada, 17 reseñas por habitación de hotel, y tan solo 2 reseñas por habitaciones compartidas. Se podría considerar que la cantidad de reseñas, en general, son escasas, sobre todo las de las habitaciones de hotel y en las habitaciones compartidas. Para fomentar el aumento de reseñas, se propone crear un sistema de recompensas en el que cada *host* puede adherirse. Este sistema consistiría en que si un huésped ha realizado una reseña en su anterior estancia, a la próxima reserva del mismo alojamiento, recibirá un descuento del 5%. Este descuento se repartiría al 50% entre el *host* y los gastos de comisión de Airbnb. Esto conllevaría otra ventaja para el *host*, y esta sería que este estaría listado como "*Host 5%*", un nuevo filtro que indicaría que este *host* acepta el descuento del 5%, por lo que estaría teniendo mayor visibilidad y ventaja competitiva respecto al resto de *hosts* que no estuvieran adscritos a la promoción.

En cuanto a las reseñas, eliminando los hoteles que, al tener servicio de limpieza no suele haber problema, las reseñas de limpieza suelen ser las más bajas. Esto, muchas veces, sucede debido a que entre huésped y huésped no hay mucho tiempo y se realiza una tarea incompleta de limpieza. Para solucionar este problema convendría, y sobre en tiempos de COVID-19, que Airbnb tomara decisiones más estrictas respecto a la limpieza y aumentara la media mínima que debe tener un alquiler en la reseña de limpieza. De esta manera se crearía una tabla de penalizaciones donde mostrase los costes que tendría. Actualmente Airbnb cobra un 3% de comisión a cada *host* (Comisión Airbnb, 2021), por lo que se propondría la Tabla 38 como mejora

Tabla 38: Comisión aplicada dependiendo de los valores medios de limpieza . Fuente: Elaboración propia

Valor medio reseña limpieza	Comisión
5 a 4.01	2%
4 a 3.01	5%
3 a 2.01	10%
2 a 0	Expulsión

De esta manera, los *host* se tomarían más en serio la limpieza y se esforzarían en llegar al nivel más alto para obtener el beneficio de pagar un 1% menos de comisión.

3.5.2 En los *hosts*

Uno de los requisitos para que funcione la plataforma es la correcta comunicación entre el huésped y el *host*. Para ello, convendría fomentar a los *hosts* a responder a todo aquel huésped que le envíe un mensaje. Al aumentar el ratio de respuesta de los *hosts*, se estaría aumentando también las reservas en la plataforma. Para hacer esto realidad, se debería penalizar a aquellos *hosts* que estén por debajo del 90% en el ratio de respuesta, y premiar a los que estén por encima. Esto podría hacerse creando un *ranking* de usuarios donde el número 1, sea el mejor y

el 6024, el peor. Este *ranking* se actualizaría todas las semanas y comprobaría si cada uno de los *hosts* está por encima o por debajo del 90% de ratio de respuesta.

Al comprobar que los *hosts* que son *superhost* tienen una mejor calificación en todas las reseñas y un mayor volumen en la cantidad de reseñas, convendría, asimismo, fomentar a los *hosts* que se esforzaran en alcanzar a los requisitos mínimos que se requiere desde Airbnb para llegar a ser *superhost*. En este caso es, según el artículo “*Cómo convertirse en Superanfitrión - Centro de ayuda*”(s. f.), para ser *superhost* los requisitos son los siguientes: (i) haber tenido más de 10 estancias con 100 noches como mínimo; (ii) un 90 % ó más de ratio de respuesta; (iii) un 4,8 en las reseñas generales; y (iv) menos de un 1% en cancelaciones. Esto podría fomentarse controlando que *hosts* cumplen 2 de los 4 requisitos y mandarles información sobre los requisitos que todavía no cumplen a modo de recordatorio motivador. Esto se debe a, que parte de los *hosts* en Airbnb desconocen de la existencia de esta forma de *host*. Por otro lado, al resto de *superhost*, sería muy positivo mandarles información trimestral sobre todo lo que le falta para llegar a ser *superhost*, junto con una guía de buenas prácticas de los pasos que debe de seguir para ser considerados superanfitriones.

3.5.3 Por género

Con el análisis de la cantidad de licencias que tenía cada género se ha llegado a la conclusión que, aunque había diferencias pequeñas entre hombres y mujeres, no eran tan grandes como para poder afirmar que los hombres tenían más licencias porcentualmente que las mujeres. Aun así, destaca el bajo número de licencias (cerca de un 40% tienen licencias, mientras que los grupos tienen más de un 60%). Tener tan pocas licencias es un aspecto muy negativo, ya que en Valencia es obligatorio tener licencia para tener una vivienda turística en alquiler, según *Valencia - Centro de ayuda de Airbnb*, (s. f.), por lo que si se permite desde Airbnb (a pesar de que se obtenga rendimiento económico) que propietarios alquilen casas sin licencias, el ayuntamiento puede poner pegas a la plataforma para ejercer su función en esta ciudad. Esto se debe a que, normalmente, este tipo de gestiones suelen ser lentas y para todos aquellos *hosts* que nunca hayan hecho una gestión de este tipo pueda ser algo complicada de realizar, lo cual genera una gran barrera. Para fomentar que los *hosts* tengan licencia, una de las propuestas de mejora es enviar un correo electrónico a todos los *hosts*, notificándoles que quién no tenga licencia, dejará de ser *superhost* y que, a partir del año 2023, necesitarán una licencia en regla para poder listar sus propiedades en Airbnb. Para ello, se enviará una guía con los pasos a seguir para conseguir la licencia. Dicha licencia conlleva de un pequeño pago que será descontado de las comisiones de Airbnb por lo que al *hosts* le saldrá gratis obtener la licencia.

En cuanto a las propiedades que tiene cada *host*, se ha comprobado que los *hosts* individuales tienen menos propiedades, como era de esperar, que los *hosts* que son grupos empresariales como hoteles o inmobiliarias. Los *hosts*, que son individuales, tienen de media 1,15 casas, es decir, que la mayoría de los *hosts* tienen 1 casa listada en la plataforma. Airbnb fue una *startup* y tiene mentalidad emprendedora por lo que sería positivo para la empresa seguir inculcando esos valores y fomentando a sus actuales *hosts* a reinvertir las ganancias mejorando sus propiedades actuales o invirtiendo en nuevas. Esto es un tema delicado ya que se introduce el tema económico de por medio pero con una correcta planificación y formación podría ser llevado a cabo. Para ello, Airbnb podría ofertar cursos *online* de formación gratuitos para todos aquellos *superhost*. Estos cursos versarán sobre finanzas y planificación de inversiones inmuebles.

3.5.4 Por barrios

En primer lugar, hay que destacar la cantidad de camas que ponen los *hosts* por habitación. Como se ha analizado anteriormente, hay apartamentos de 1 habitación que los alquilan con 9 camas; o apartamentos con 2 habitaciones, que tienen 22 camas, es decir, a una media de 11 camas por habitación. Esto es algo abusivo que desde Airbnb habría que limitar, ya que seguramente no cumpla con la normativa de habitabilidad, ya que según especifica *Ingeniería*, (s. f.), por cada 3 personas debe haber una superficie mayor a 12 m².

Para controlar este aspecto, una propuesta de mejora sería limitar la cantidad de camas por habitación, haciendo una restricción de 3 camas por habitación más el sofá/sofa-cama; estableciendo, de esta manera, un máximo de 5 personas en un apartamento de 1 habitación.

A pesar de que la disponibilidad media ha bajado durante el último mes en relación con el último año, sería muy positivo si se pudiera reducir todavía más. Esto se podría llevar a cabo realizando un aumento de publicidad para llegar a más personas. Airbnb tenía previsto gastar 800 millones de dólares en publicidad según Marketing (2020).

Por otro lado, analizando los barrios con una mayor disponibilidad media, se podría ofrecer desde la plataforma, los medios de transporte para ir a otras zonas de la ciudad y las actividades turísticas que tienen esos barrios, para, de esta manera, fomentar la reserva en esas zonas. Actualmente ya hay plataformas que recapitulan lo que se puede hacer en estas zonas, por ejemplo TripAdvisor tiene una guía de las cosas que se pueden hacer en Benicalap, uno de los barrios con mayor disponibilidad de Valencia (*Los 10 mejores cosas que hacer cerca de Parque de Benicalap, Valencia - Tripadvisor*, s. f.).

Por otro lado y teniendo en cuenta lo que hacen empresas como Uber, que al detectar una mayor demanda aumentan el precio para así obtener siempre un beneficio máximo, en Airbnb se podría implementar una estrategia similar. En el momento que queden pocos días disponibles, la propuesta englobaría enviar un mensaje al *host* aconsejándole que si sube el precio cierta cantidad, podría obtener un beneficio mayor. En este sentido, se podría desarrollar un modelo matemático, cuya variable de decisión fuese el precio, teniendo en cuenta parámetros de disponibilidad, fechas y zonas, entre otros.

En cuanto a los mapas de distribución de los barrios con mayor disponibilidad se observa que son todos aquellos barrios que no están situados ni en el centro de la ciudad ni en la playa. Para tratar de mejora esta situación e intentar igualar las disponibilidades, Airbnb podría desarrollar y proporcionar a los *hosts* una serie de guías que muestren los beneficios de reservar en esos barrios. Estos beneficios pueden ser los de tener una mayor tranquilidad, mejor aire o menor contaminación.

Finalmente, viendo los barrios agrupados por zonas (zona playa, zona centro y el resto de las zonas), se puede observar cómo los alquileres en las zonas de playa tienen menor disponibilidad media que los alquileres en la zona del centro. Una recomendación que podría hacer la plataforma de Airbnb a los *hosts* con alquileres en la playa sería aumentar los precios. Esta recomendación viene fundamentada en que los alquileres en la zona de playa tienen la disponibilidad de los alquileres en la zona del centro, pero con los precios por cama por noche de los barrios del extra-radio.



3.5.5 En las reseñas

Las reseñas son muy útiles para conocer que está haciendo mal cada *host*, pero sobre todo son útiles para ver cuáles son los idiomas más utilizados en la comunicación *host*-inquilino. En este caso, el análisis ha mostrado, en 3 de los últimos 4 años, el inglés como el idioma más hablado. Seguido del inglés, el idioma más utilizado ha sido el español y después el francés, alemán, holandés e italiano. Con esto no se quiere p a la conclusión de que un *host* tenga que aprender necesariamente todos estos idiomas, sino que sería conveniente que, como mínimo, sepa hablar inglés y entender algún otro idioma más. Esto hará que haya una mejor comunicación, con lo cual desde Airbnb, se podría realizar publicidad en este sentido e incluso ofrecer cursos relacionados con idiomas.

4. Conclusiones

Desde que nació el intercambio económico, este ha experimentado muchas revoluciones. La más reciente ha sido la economía colaborativa, la cual ha sabido, gracias a los avances tecnológicos, crear valor, y hacerlo llegar a la sociedad.

El sector del alquiler de viviendas turísticas es de los que más se han beneficiado de este nuevo paradigma colaborativo. Es por ello, que el presente TFG ha estudiado uno de los casos más populares de la gestión de alojamientos colaborativos: la plataforma Airbnb. Para ello se han obtenido datos de Airbnb para realizar un análisis del tipo de alquileres, hosts, aspectos de género, barrios de Valencia y reseñas, para tratar de identificar debilidades y puntos fuertes y, de esta manera, proponer una serie de acciones de mejora que puedan repercutir en los beneficios obtenidos por los *hosts*, así como por Airbnb y en los usuarios finales que alquilan dichos alojamientos.

Del análisis de los resultados, destacar que como es de esperar, la mayoría de los alquileres son de casas enteras, ya que normalmente cuando se viaja se quiere tranquilidad y seguridad y eso normalmente lo da una casa entera. Esto hace que los alquileres enteros tengan los valores más elevados, desde la cantidad de licencias, que tienen una mayor cantidad respecto a los otros tipos de alquileres como pueden ser las habitaciones privadas o compartidas, pero también tienen una mayor cantidad de reservas, una mayor cantidad de *host*, y de esos *host* tienen una mayor cantidad de *superhost* que el resto. En cuanto a los géneros, tanto el masculino como el femenino no presentan prácticamente diferencias. En lo que respecta a los barrios queda muy claro cuáles son los barrios o zonas más queridas por los huéspedes para realizar las reservas. Valencia al tener un paseo marítimo muy grande y con mucha vida tanto diaria como nocturna es mucha la gente que quiere hospedarse ahí, y un centro muy atractivo. Es por ello que son estas dos zonas, las que mayor cantidad de reservas tienen. Finalmente, en cuanto a las reseñas, como era de esperar la mayoría de *host* hablan en inglés llevándose así este idioma casi la mitad de las reseñas.

De la detección y propuesta de mejoras cabría destacar los esfuerzos que se debe hacer en Airbnb para paliar los efectos negativos de la concentración de casas en alquiler en puntos muy marcados. Por otro lado, y en lo que se refiere a las reseñas convendría crear un sistema de recompensas que alentase al huésped a realizar la reseña que muchas veces no se realiza. Continuando con las reseñas, este caso la de limpieza y más debido a la pandemia, desde Airbnb se tiene que tomar decisiones drásticas que controlen al máximo eso, haciendo que al *host* le cueste mayor cantidad de dinero tener un piso listado en Airbnb si no cumple unos mínimos en la limpieza. Por otro lado, es muy conveniente aumentar la cantidad de licencias que tienen los apartamentos, ya que no es un requisito voluntario, sino obligatorio. A fecha de realización de este trabajo, con los datos usados, solamente un 40% tienen licencias. Aumentar esto haría que la administración pública viera que la plataforma está volcada con mantener una ciudad próspera y con un cierto orden. Aumentar la cantidad de licencias sería posible si desde Airbnb se concienciara de ello y se pusiera una fecha límite para tener licencia y no perder los beneficios como los que otorga ser *superhost*. Además de ello desde Airbnb se propone pagar las tasas de la obtención de la licencia mediante las comisiones de alquiler.

Una de las principales limitaciones del presente TFG es la falta de información económica para poder analizar en detalle la viabilidad de las propuestas de mejora. Como líneas futuras de

trabajo, se podrían cuantificar las propuestas en valores monetarios, sin embargo se precisaría de cómo afectaría económicamente su implementación en Airbnb, en los *hosts* y en los usuarios.

En cuanto a las regulaciones, es positivo que haya autonomía por parte de los *hosts* en alquilar los apartamentos que se desee, pero es más beneficioso para la sociedad en conjunto que haya cierta regulación que controle que no haya abusos en cuanto a los alquileres. Estos abusos pueden ser por ejemplo los de poner más camas en una casa que las que permite la ley.

Por otro lado, se debe fomentar una pequeña profesionalidad a la hora de gestionar pisos, ya que, en muchas ocasiones, se realiza de manera descontrolada y sin tener conocimientos financieros.

En cuanto a las líneas futuras del trabajo sería muy interesante hacer un análisis horizontal, es decir, un análisis con datos de varios años para poder estudiar las tendencias de cada barrio y en conjunto la ciudad de Valencia. De esta manera se podría analizar si con los nuevos cambios propuestos se estaría yendo en buena dirección. Por otro lado, el estudio de la tendencia posibilitaría otras mejoras distintas a las vistas en este análisis.

Finalmente destacar que Airbnb está relacionado con el concepto de economía colaborativa que es modelo económico en el que, fundamentalmente, los usuarios se aprovechan de las nuevas tecnologías para prestar, comprar, vender, compartir o alquilar bienes y servicios, y por ello, su alineamiento con el objetivo de desarrollo sostenible 8, 11.

Referencias

5 ejemplos de uso real de data analytics. (2018, 15 junio). Máster en Data Science. Recuperado 22 de julio de 2022, de <https://www.master-data-scientist.com/ejemplos-master-en-data-analytics/>

Los 10 mejores cosas que hacer cerca de Parque de Benicalap, Valencia - Tripadvisor. (s. f.). Tripadvisor. Recuperado 23 de julio de 2022, de https://www.tripadvisor.es/AttractionsNear-g187529-d243201-Parque_de_Benicalap-Valencia_Province_of_Valencia_Valencian_Country.html

Las 10 mejores herramientas big data. (2021, 22 octubre). Afi Escuela de Finanzas. Recuperado 21 de julio de 2022, de <https://www.afiescueladefinanzas.es/las-10-herramientas-big-data-mas-importantes>

Análisis Kilpatrick [Comentario sobre el artículo “EL MÉTODO DE PROYECTOS DE KILPATRICK”]. (1910, 1 enero). <http://temasdeeducacin.blogspot.com/2008/04/el-mtodo-de-proyectos-de-kilpatrick.html>. <http://temasdeeducacin.blogspot.com/2008/04/el-mtodo-de-proyectos-de-kilpatrick.html>

Apéndice B. Idiomas — Lista de códigos ISO 639. (s. f.). <https://omegat.sourceforge.io/>. Recuperado 23 de julio de 2022, de <https://omegat.sourceforge.io/manual-standard/es/appendix.languages.html>

Comisión airbnb. (2021, 14 septiembre). Airbnb. Recuperado 23 de julio de 2022, de <https://www.airbnb.es/resources/hosting-homes/a/how-much-does-airbnb-charge-hosts-288#:~:text=%C2%BFCu%C3%A1nto%20cobra%20Airbnb%3F>

Cómo convertirse en Superanfitrión - Centro de ayuda de. (s. f.). Airbnb. Recuperado 23 de julio de 2022, de <https://www.airbnb.es/help/article/829/c%C3%B3mo-convertirse-en-superanfitri%C3%B3n>

Cuatro áreas en las que puede emplearse el análisis de datos. (2020, 7 enero).

threepoints.com. Recuperado 21 de julio de 2022, de <https://www.threepoints.com/blog/data-analytics-para-principiantes#:~:text=Seg%C3%BAn%20la%20empresa%20de%20investigaci%C3%B3n%20de%20mercados%20Gartner%2C%20existen%20cuatro%20%C3%A1reas%20en%20las%20que%20puede%20emplearse%20el%20an%C3%A1lisis%20de%20datos>

Data analytics para principiantes. (2020, 7 enero). Threepoints. Recuperado 21 de julio de

2022, de <https://www.threepoints.com/blog/data-analytics-para-principiantes>

Get the data. (2022). [Http://insideairbnb.Com/](http://insideairbnb.com/). Recuperado 22 de junio de 2022, de

<http://insideairbnb.com/get-the-data/>

Ingeniería, A. E. (s. f.). *¿Cuántas personas pueden vivir en una vivienda?* sinluz.com.

Recuperado 23 de julio de 2022, de

<http://www.sinluz.com/es/servicios/arquitectura/cedula-de-habitabilidad/306-cuantas-personas-pueden-vivir-en-una-vivienda>

Marketind Directo. (2020, 30 marzo). *Inversión de airbnb.* Marketing Directo. Recuperado 23

de julio de 2022, de <https://www.marketingdirecto.com/anunciantes-general/anunciantes/airbnb-frena-su-inversion-en-marketing-valorada-en-la-friolera-cantidad-de-800-millones-de-dolares-anuales>

Valencia - Centro de ayuda de Airbnb. (s. f.). Airbnb. Recuperado 23 de julio de 2022, de

<https://www.airbnb.es/help/article/1208/valencia#:~:text=de%20comercializaci%C3%B3n%20tur%C3%ADstica.-,Registro,-Es%20obligatorio%20inscribir>

