

Inversiones inmobiliarias para alquiler vacacional - Paris, Île-de-France, France

Lina María Gómez Mesa, Universidad de los Andes, l.gomez1@uniandes.edu.co

1. Introducción

El presente informe tiene como objetivo, ayudar a individuos o empresas que buscan invertir en propiedades para alquilar a través de la plataforma AirBnb. Con la creciente popularidad de este modelo de negocio, es cada vez más difícil tomar buenas decisiones respecto a los mejores sectores de la ciudad, tipos de propiedad, entre otros factores, para lograr una rentabilidad adecuada a lo largo del tiempo. Una vez que la propiedad se encuentra listada en la plataforma, asegurar una alta tasa de ocupación estableciendo un precio por noche justo y un buen sistema de atención a los usuarios es vital para el éxito del negocio.

Bajo este supuesto, y como equipo consultor experto en el negocio, se decidió tomar como base de datos Inside AirbnB que provee propiedades listadas en AirBnb para una ciudad específica. Por cada propiedad se encuentran atributos como ubicación, detalles del host, tipo de propiedad, precio, disponibilidad futura, calificaciones, entre otros. Paris, es una de las ciudades con mayor número de AirbnB en el mundo presentando un total de **95461 registros** de AirbnB a la fecha del **6 de septiembre del 2024**. En Francia, por ejemplo, sus ejecutivos informan que sus visitantes contribuyen significativamente a la industria turística de Francia, empresas y hogares locales. Según un Informe AirbnB, la empresa ha contribuido al aumento del turismo en Francia. De hecho, informan que el 23% de los huéspedes de AirbnB dijeron que no hubieran ido de viaje o no se hubieran quedado tanto tiempo sin AirbnB [1]. Es más, se reporta que 1/3 de los usuarios visita Paris [2].

Es por lo anterior, que Paris se considera uno de los principales epicentros en los que los inversionistas pueden invertir y obtener utilidades hacia el futuro.

2. Entendimiento inicial de los datos

Inicialmente, se tomó el conjunto de datos y se realizó un análisis preliminar de las columnas encontradas. En total se obtuvieron **95461 filas** (propiedades) y **75 columnas** en el dataframe de listings. Se encontraron 40 columnas numéricas y 35 columnas categóricas. Además, se hizo uso de la librería de Python de *pandas_profiling* que realiza un análisis detallado de las principales propiedades encontrados en el dataset. A continuación, en la figura 1 se muestra un resumen detallado de los principales findings encontrados. Pandas profiling difirió en el tipo de variables clasificándolas de forma más detallada incluyendo 37 numéricas, 18 tipo texto, 10 categóricas, 3 tipo fecha, 5 booleanas y 2 sin tipo. Por otro lado, mostró que el 13.1% de las filas tenían información faltante.

Overview

Alerts 95

Reproduction

Dataset statistics

Number of variables	75
Number of observations	95461
Missing cells	937829
Missing cells (%)	13.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	54.6 MiB
Average record size in memory	600.0 B

Variable types

Numeric	37
Text	18
Categorical	10
DateTime	3
Boolean	5
Unsupported	2

Figura 1. Resumen general de pandas profiling sobre el dataset de Paris

A partir de la información obtenida de *pandas profiling* y con la ayuda del diccionario de datos y las definiciones se encontraron como relevantes las siguientes variables para los inversores:

Tabla 1. Columnas seleccionadas preliminarmente

Atributo	Definición
Room_type	[Entire home/apt,Private room,Shared room,Hotel]
Price	Daily price in local currency
Availability_365	The availability of the listing x days in the future as determined by the calendar.
Number of reviews	The number of reviews the listing has
Neighbourhood_cleansed	The neighbourhood as geocoded using the latitude and longitude against neighborhoods.
Accommodates	The maximum capacity of the listing.
License	The licence permit, registration number.
Property_type	Self-selected property type. Hotels and Bed and Breakfasts are described as such by their hosts in this field.
Amenities	features or services that enhance the comfort, convenience, or enjoyment of a space, facility, or community
Review_scores_rating	Score given to the place
Review_scores_location	Score given to the location
Minimum_nights	Minimum number of night stay for the listing (calendar rules may be different)

Estas son variables que de acuerdo con previos estudios hechos sobre el mismo dataset, pueden ayudar a predecir el precio de las acomodaciones [3]. Además, teniendo en cuenta que el objetivo de los inversores es maximizar sus ingresos se consideran relevantes con el fin de poder determinar el precio de las propiedades listadas en Airbnb para París. Además, se realizó un análisis preliminar de completitud sobre estas variables, encontrando que el precio presentaba 32.72% de sus valores nulos, Review_scores_location el 28.45%, Review_scores_rating el 28.43% y licencia el 20.37%. Dado el supuesto anteriormente mencionado sobre el precio como variable dependiente, si bien esta presenta valores nulos, se decidió dejar dada la relevancia para los inversores. El resto de las variables sí fueron descartadas ya que presentaban valores nulos.

Por otra parte, siguiendo estudios anteriores se observó que al realizar una regresión lineal multilínea, el tipo de cuarto (room-type) adquiría relevancia a la hora de determinar los precios en conjunto con el número de huéspedes (accommodates) [4]. Por otro lado, con el fin de calcular columnas derivadas como los posibles ingresos se consideró la columna de availability_365 que es inversa a la tasa de ocupación. Finalmente, se consideró la variable barrios ya que, esta influye en la seguridad y qué tan cerca se encuentra de lugares turísticos.

A continuación, se realizó el análisis univariado para las cinco variables seleccionadas. Se encontró que el precio promedio es de \$171 euros y la mediana es \$148 cuando se quitan los outliers (5770 filas con valor superior a 485.5 euros). El 89.32% de las propiedades son Entire home/apt. El barrio con más Airbnb es Buttes-Montmartre (11.03%) y la mayoría tienen una disponibilidad media de 127 días y una mediana de 84 días. Finalmente, la cantidad de miembros más repetido por propiedad es de 2 miembros por Airbnb.

3. Estrategia de análisis

Como estrategia de análisis para encontrar los mejores Airbnb para invertir, se considerará una estrategia de tres pasos: **1)** análisis descriptivo, **2)** análisis visual (gráficos) y **3)** pruebas de hipótesis. Estos se describen a continuación y se encuentran resumidos en la figura 2:

- A. Análisis descriptivo: Se calcularán estadísticos descriptivos como media, mediana, desviación estándar y cuartiles para variables numéricas clave como el precio, la capacidad de alojamiento (*accommodates*) y la disponibilidad anual (*availability_365*). Para las variables categóricas como el tipo de habitación y el tipo de propiedad, se examinará la moda y las frecuencias relativas. Esta exploración inicial permitirá comprender la distribución de los datos, identificar

valores típicos y atípicos, y detectar posibles sesgos o tendencias en el mercado. Además, se analizará completitud, presencia de outliers y consistencia para las variables.

- B. **Análisis visual:** Se emplearán gráficos de caja (boxplots) y/o gráficos de barras apilados para visualizar la distribución del precio en función del tipo de habitación y el tipo de propiedad. Esta técnica permitirá comparar visualmente los precios entre diferentes categorías, revelando diferencias en los rangos de precios y la dispersión de los datos, lo que ayudará a identificar los tipos de alojamiento más lucrativos dependiendo si la variable es numérica o categórica. Finalmente, se construirán mapas de calor de correlaciones para examinar las relaciones entre todas las variables numéricas, incluyendo el precio, la disponibilidad y la capacidad de alojamiento.
- C. **Pruebas de hipótesis:** Se realizarán pruebas de hipótesis, como análisis de varianza (ANOVA) o Kruskal wallis y pruebas de pearson, para evaluar si existen diferencias significativas en el precio o ingresos según el tipo de habitación, el tipo de propiedad y la presencia de reseñas. Estas pruebas estadísticas proporcionarán evidencia sólida para respaldar las recomendaciones de inversión.

Plan de estrategia

Con el fin de recomendar las mejores propiedades en la ciudad de París, se recomienda seguir una estrategia de tres pasos: análisis descriptivo, data visualizations y pruebas de hipótesis

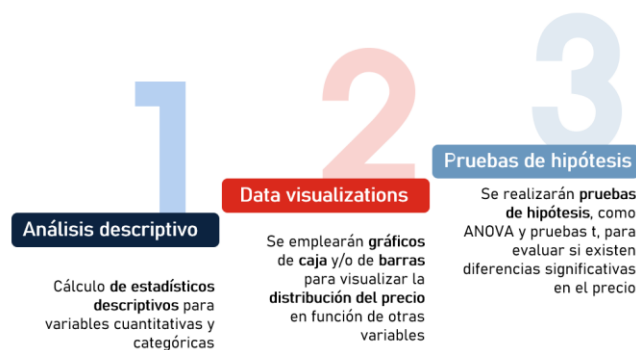


Figura 2. Pasos del plan de estrategia analítico

4. Desarrollo de la estrategia

En primer lugar, se realizó el análisis con las variables numéricas tanto para el precio, la disponibilidad en 365 días al año y la cantidad de personas admitidas por propiedad.

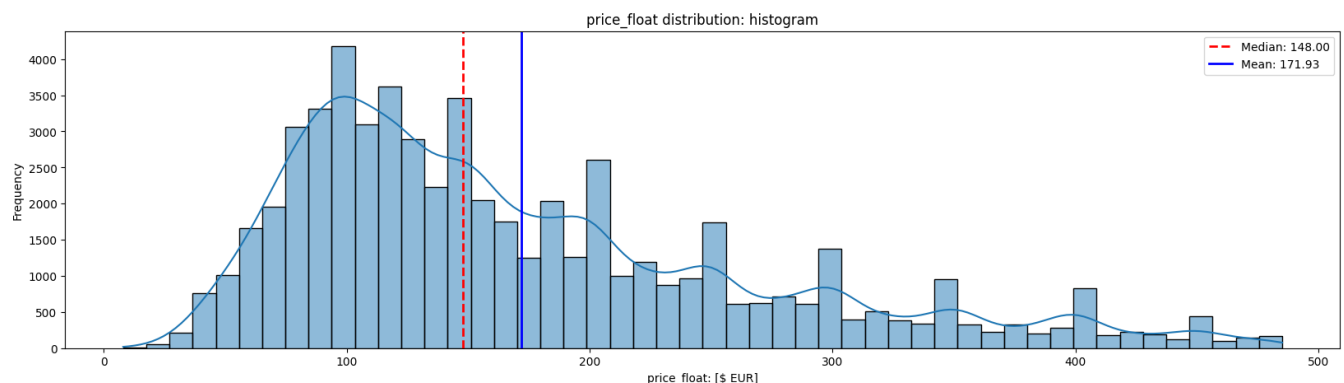


Figura 3. Distribución del precio

Las propiedades de Airbnb en París tienen un precio medio de aproximadamente 256 euros, aunque con una gran variabilidad, evidenciada por una desviación estándar de 522 euros. La presencia de 32.72% de valores nulos en la columna de precios indica que un número significativo de registros no tienen precio/noche, lo cual puede afectar la precisión en el análisis para los inversionistas. Además, el análisis de los valores atípicos muestra que un 6.04% de los precios se consideran extremos, lo que sugiere que hay apartamentos con precios significativamente más altos. La distribución también presenta una fuerte asimetría positiva (skewness de 17.25) y alta kurtosis (kurtosis de 507.77), lo que indica que la mayoría de los apartamentos tienen precios relativamente bajos, mientras que unos pocos tienen precios extremadamente altos. Si se quitan los outliers se presenta la media y mediana que se muestran en la figura 3. En este caso, se propuso eliminar los valores nulos y quitar los outliers.

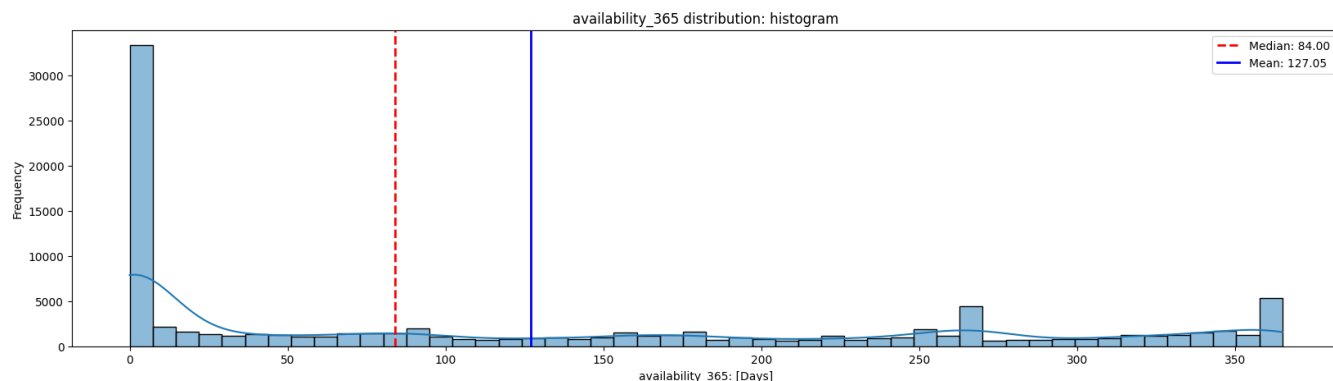


Figura 4. Distribución de la disponibilidad en 365 días del año

Las propiedades en París presentan una disponibilidad promedio de 127 días. La desviación estándar es de 130.79 días, lo que indica una variabilidad considerable en la disponibilidad de los apartamentos. El valor mínimo es 0, lo que sugiere que algunos apartamentos no están disponibles en absoluto durante el año, mientras que el máximo es de 365 días, lo que indica que hay propiedades disponibles todo el año. La mediana, ubicada en 84 días, señala que la mayoría de los apartamentos tienen una disponibilidad relativamente baja. Además, no se presentan valores nulos en esta columna, lo que garantiza que todos los registros cuentan con información sobre la disponibilidad. En cuanto a los outliers, no se han identificado valores atípicos, lo que indica que la mayoría de los datos se ajustan a un rango razonable. La asimetría (skewness de 0.52) sugiere una ligera inclinación positiva en la distribución, lo que indica que hay un número considerable de apartamentos con disponibilidad limitada. La kurtosis negativa (-1.26) sugiere que la distribución es más plana que la normal, lo que implica que hay menos frecuencias extremas en comparación con una distribución normal. Para esta columna no se proponen cambios.

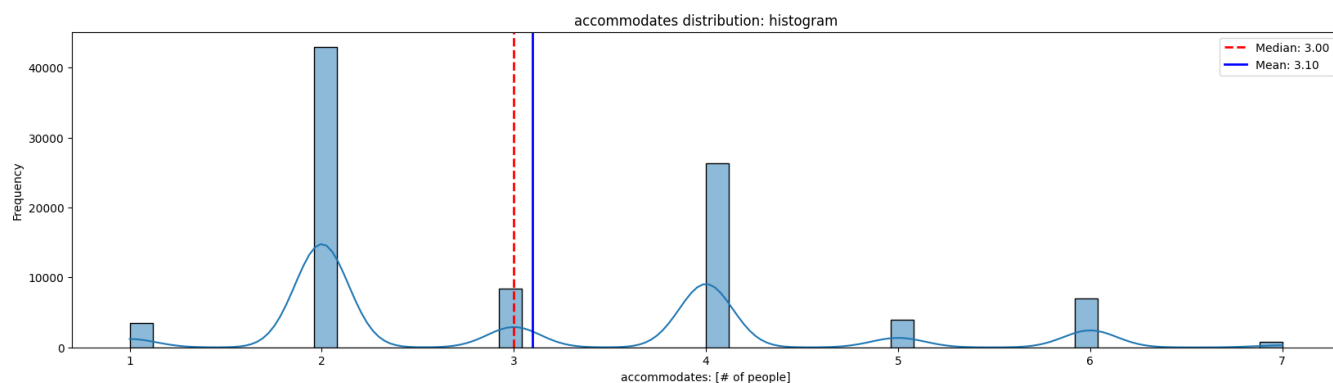


Figura 5. Distribución de la cantidad de personas admitidas por propiedad

El promedio es de aproximadamente 3 personas, lo que sugiere que la mayoría de los apartamentos están diseñados para grupos pequeños. La desviación estándar es de 1.66, por lo que se presenta cierta variabilidad en la capacidad de alojamiento. Los valores en esta columna oscilan entre un mínimo de 1 y un máximo de 16 huéspedes. La mediana es de 3, lo que refuerza la idea de que muchos apartamentos están orientados a grupos reducidos, ya que el 50% de los registros tienen una capacidad de 3 o menos personas. No hay valores nulos en esta columna, lo que garantiza que todos los registros tienen información sobre la capacidad de alojamiento. La asimetría (skewness de 1.59) indica una inclinación positiva en la distribución, lo que sugiere que hay una mayor concentración de apartamentos con capacidades más bajas y algunos pocos con capacidades significativamente más altas. Si se presentan outliers por lo que estos serán removidos.

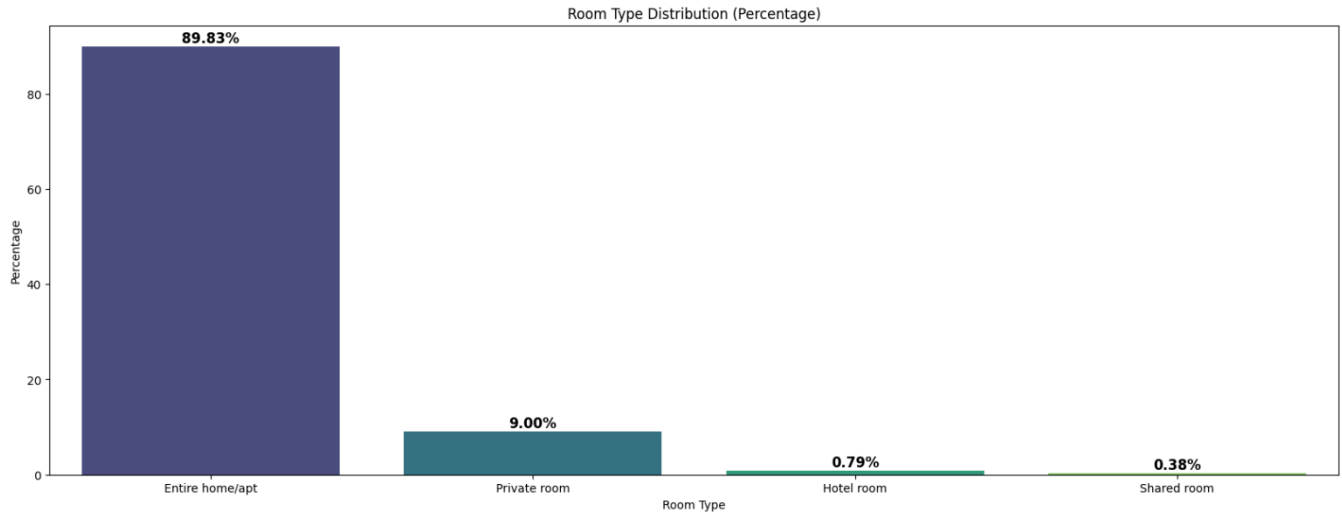


Figura 6. Distribución de la cantidad de personas por el tipo de habitación

Esta variable se compone de cuatro categorías únicas: "Entire home/apt", "Private room", "Hotel room" y "Shared room". La categoría más frecuente es "Entire home/apt", que representa aproximadamente el 90% de los registros, seguida de "Private room" con alrededor del 9%. Las categorías "Hotel room" y "Shared room" son significativamente menos comunes, representando menos del 1% cada una. Esto se evidencia en la figura 6.

Por otro lado, la variable de los barrios cuenta con 20 barrios únicos, siendo "Buttes-Montmartre" el más común, con 6,185 registros, lo que representa aproximadamente el 10.58% del total. Otros barrios destacados son "Popincourt" (8.60%), "Vaugirard" (8.11%) y "Batignolles-Monceau" (6.92%), que también tienen una proporción significativa de registros. Esta variabilidad sugiere que las preferencias de los usuarios pueden estar influenciadas por la ubicación en la ciudad, y sería interesante explorar si hay diferencias en precios o tipos de alojamientos en función de los barrios. Esto se evidencia en la figura 7.

Por otro lado, con estas variables, se pueden calcular nuevas columnas derivadas como la tasa de ocupación, y los ingresos. Las formulas utilizadas se presentan a continuación:

$$(1) \quad Tasa_{ocupación} = \frac{365 - availability_{365}}{365}$$

$$(2) \quad Income_{generated} = Tasa_{ocupación} * Price_{float} * 365$$

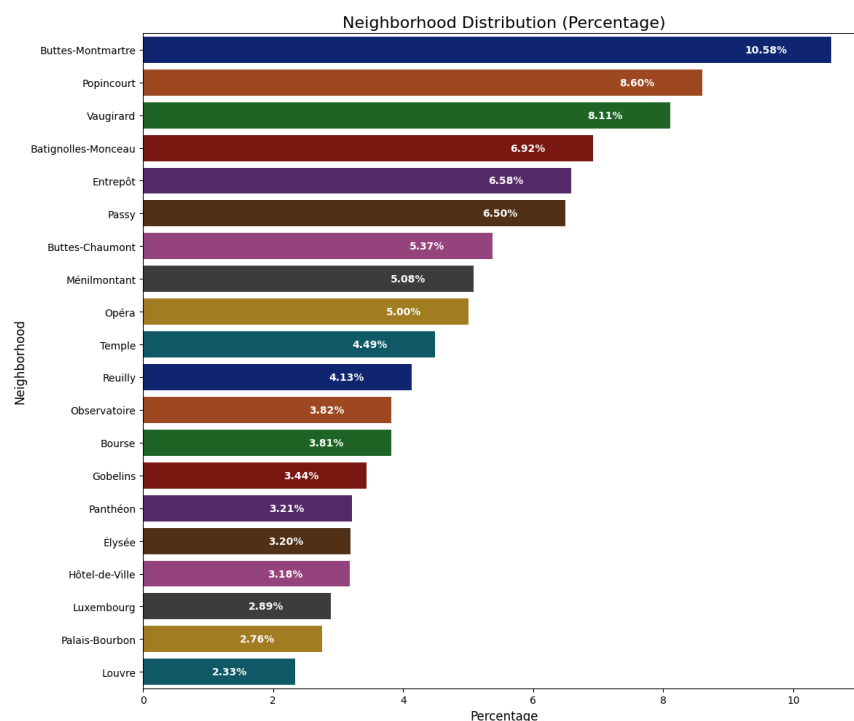


Figura 7. Distribución de la cantidad de personas por el barrio

En segundo lugar, se realizó la etapa de visualizaciones de los datos teniendo como principal variable target a observar a los ingresos generados, la cual es una combinación de availability-365 y el precio. Se observa en la figura 8 que entre las cuatro posibles tipos de habitación las que presentan mayores ingresos son Entire home/apt y Hotel room.

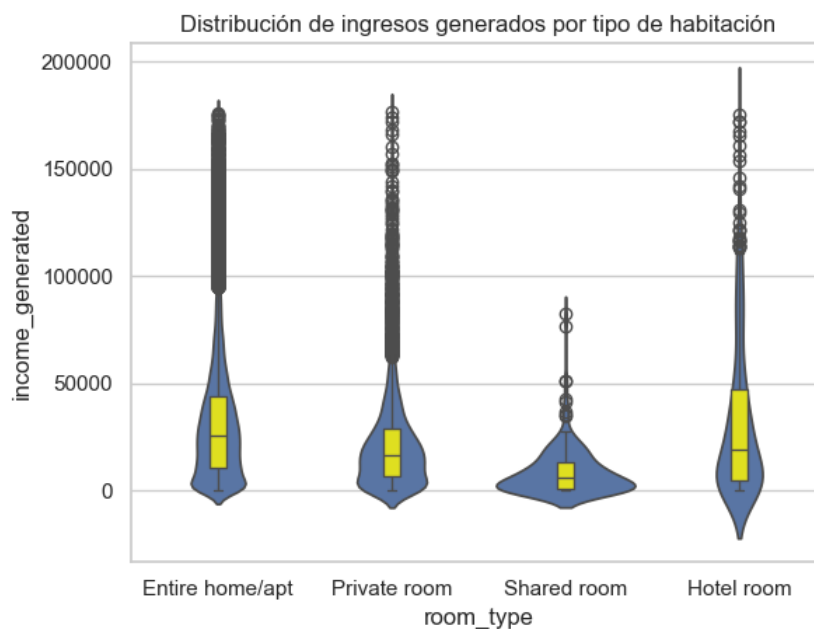


Figura 8. Distribución de los ingresos generados con el tipo de habitación

Entire home/apt tiene la mediana más alta (25,658), lo que indica que la mitad de las propiedades de este tipo generan al menos este monto. La media (31,590) también es alta, lo que sugiere que algunas propiedades generan ingresos muy superiores a la mediana. Por otro lado, Hotel room tiene una mediana menor (18,950) pero una media considerablemente mayor (32,891), lo que indica que algunas habitaciones de hotel generan ingresos muy elevados, lo que arrastra la media hacia arriba.

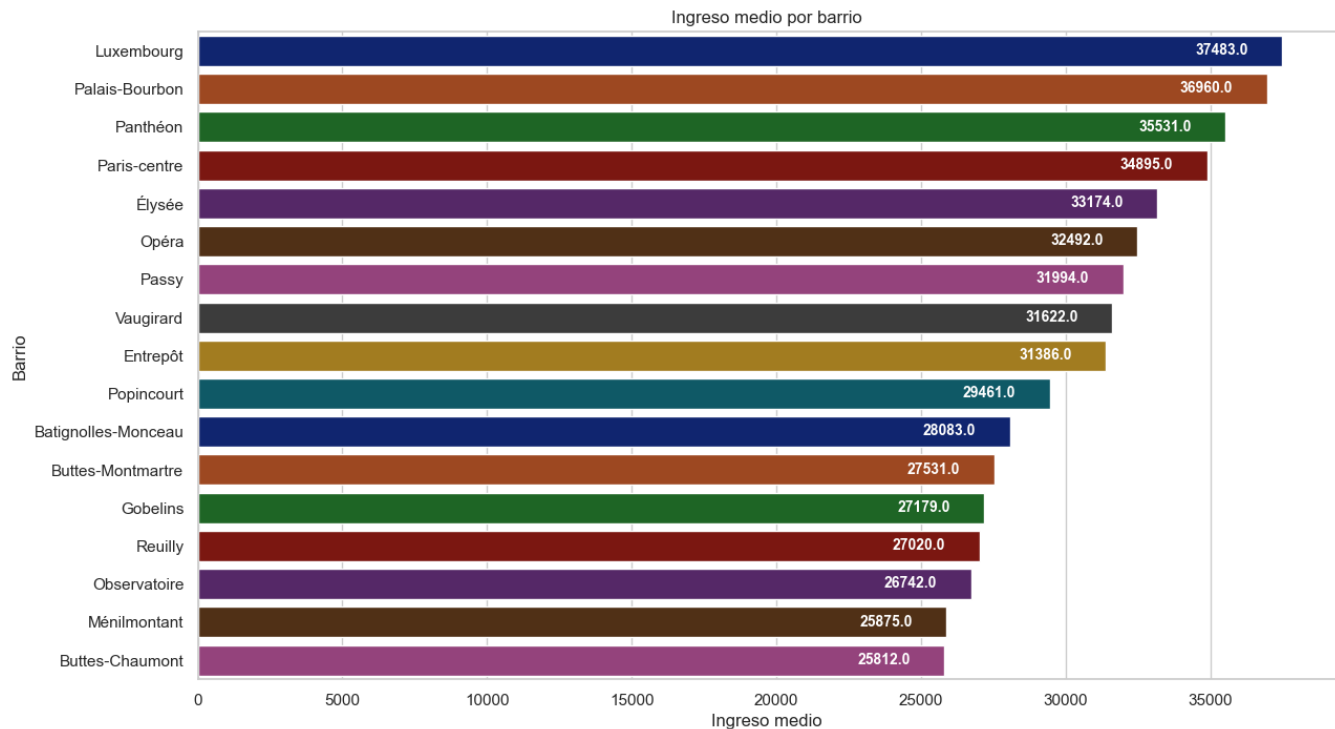


Figura 9. Distribución de los ingresos generados con el barrio

En este caso para los barrios, se observa que Luxembourg y Palais-Bourbon son los barrios con los ingresos más altos, con una mediana de alrededor de 30,000 euros y una media cercana a 37,000 euros. Por otro lado, Panthéon y Paris-centre presentan ingresos robustos, con medianas alrededor de 28,000-29,000 euros. Son buenos candidatos para inversión, ya que están ubicados en áreas centrales y populares. Finalmente, en la figura 10 se muestra que en términos de ingresos, aquellos Airbnb que reciben entre 7 y 9 personas son los que obtienen mayores ingresos.

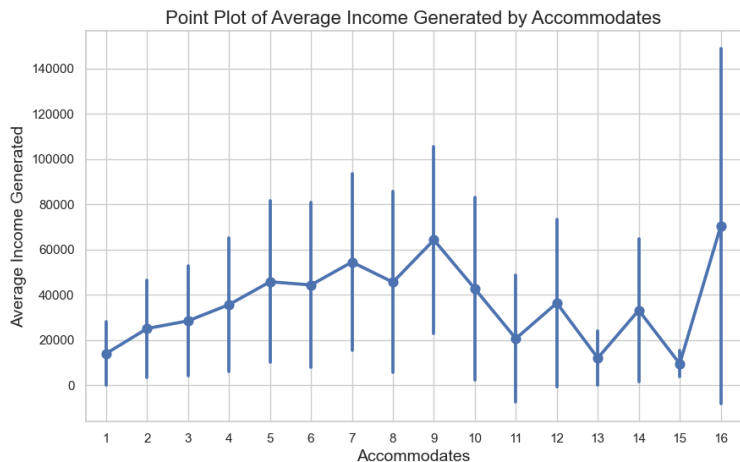


Figura 10. Distribución de los ingresos generados con el número de miembros

Finalmente, se realizó la comparación entre los barrios, el tipo de habitación y los ingresos en la que se vio que aquellos que presentan buenos ingresos son: 1)Luxembourg, Palais-Bourbon, Panthéon, Paris-centre con Entire apt y 2) Ménilmontant, Paris-Centre y Passy para hoteles.

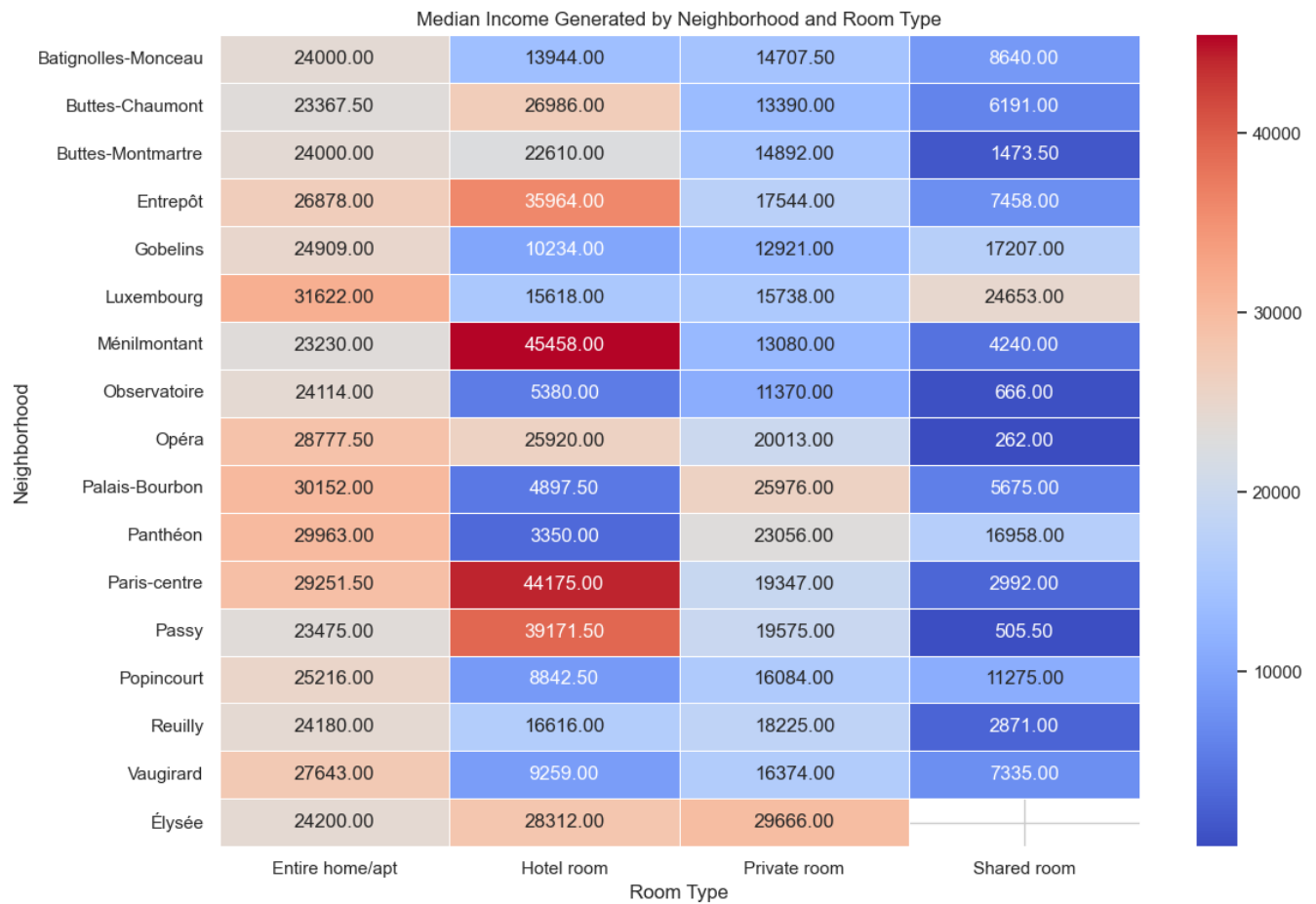


Figura 11. Distribución de los ingresos generados con el tipo de habitación y barrio

Finalmente, se realizó pruebas de hipótesis tipo Kruskal-Wallis dado que los datos no siguen una distribución normal por lo que no se hizo One-way Anova. Esta prueba de hipótesis se realizó al comparar variables categóricas vs numérica (ingreso). Dado que el valor p-value dio menor al nivel de significancia (0.05) para ambos casos, sí se presenta diferencias estadísticamente significativas en los ingresos generados entre los diferentes tipos de habitación y barrios por lo que las gráficas obtenidas sí pueden ayudar a los inversionistas a tomar mejores decisiones. Por otro lado, para la variable numérica de accommodates se utilizó el test de spearman ya que tampoco sigue una distribución normal. Este dio un valor de aproximadamente 0.36 lo que muestra una correlación débil entre el número de personas permitidas y los ingresos.

5. Conclusiones

Teniendo en cuenta los resultados obtenidos, y las figuras de la 6 a la 11, se obtuvieron las siguientes conclusiones sobre dónde invertir: 1) Luxembourg, Palais-Bourbon, Panthéon, Paris-centre cuando son Entire home/apt y 2) Ménilmontant, Paris-Centre y Passy para hoteles. En estos por tipo de *accommodate* los que presentan mejores ingresos son:

- Luxembourg/Entire apt: 5 personas (57240 eur) y 9 personas (93726 eur)
- Palais-Bourbon/Entire apt: 7 personas (53655 eur) y 8 personas (59381 eur)

- Panthéon/Entire apt: 7 personas (66440 eur) y 12 personas (75602 eur)
- Paris-centre/Entire apt: 5 personas (48000 eur) y 7 personas (42840 eur)
- Ménilmontant/Hotel: 4 personas (69408 eur)
- Paris-centre/Hotel: 1 persona (98969 eur) y 4 personas (64495 eur)
- Passy/Hotel: 1 persona (47450 eur) y 2 personas (48980 eur)

Este fue un approach top down en el que se fueron añadiendo la totalidad de las variables. Si bien, ya en el análisis de 4 variables que hay lugares con una cantidad específica de huéspedes que les va mejor, la mediana al únicamente compara la zona y el tipo de habitación, indica que esos son los barrios con un ingreso mediano en el que valdría la pena invertir. En particular, el top tres termina siendo:

1. Paris-centre/Hotel (1 persona): 98969 eur
2. Luxembourg/Entire apt (9 personas): 93726 eur
3. Panthéon/Entire apt (12 personas): 75602 eur

Nota: Las gráficas más detalladas con estos valores se pueden ver en el jupyter en el que se realiza un análisis multivariable.

6. Referencias

- [1] Cindy Yoonjoung Heo, Inès Blal, Miju Choi, what is happening in Paris? Airbnb, hotels, and the Parisian market: A case study, *Tourism Management*, Volume 70, 2019, Pages 78-88, ISSN 0261-5177
- [2] Airbnb (2015, May 1). Airbnb economic impacts in France. Retrieved from: <http://blog.airbnb.com/airbnb-economic-impacts-in-france/>
- [3] Herrgård, J., & Flöjs, J. (2023). *Driving factors behind Airbnb pricing: A multilinear regression analysis*.
- [4] Dataiku. (n.d.). *Finding the best Airbnb in Paris with data science*. Retrieved September 22, 2024, from <https://blog.dataiku.com/finding-the-best-airbnb-in-paris-with-data-science>