Nonlinear Analysis: Modelling and Control, Vol. vv, No. nn, 202553 © IE Infotep

Analisis Univariado y Multivariado*

Aris Avila, Julieth Gutierrez, Diego Alzate

Abstract

Este estudio analiza los datos de viviendas en California, explorando la relación entre variables clave como el valor mediano de la vivienda, el número total de habitaciones y el número total de dormitorios. Se realizan análisis univariados y multivariados para identificar patrones y correlaciones, utilizando visualizaciones como histogramas, diagramas de caja y gráficos de dispersión. Los resultados indican que ni el número de habitaciones ni el de dormitorios tienen una fuerte influencia en el valor de la vivienda, pero ambas variables están altamente correlacionadas entre sí.

1 Introducción

El mercado inmobiliario de California es un sector dinámico donde múltiples factores afectan el valor de las viviendas. Este análisis tiene como objetivo explorar cómo el número total de habitaciones y dormitorios pueden influir en el precio de la vivienda y determinar si estas variables están correlacionadas. Para ello, se realizan análisis descriptivos y se aplican técnicas de visualización para interpretar mejor los datos disponibles.

2 Objetivos del Análisis

Los principales objetivos de este estudio son:

- Analizar la distribución de cada variable relevante.
- Determinar la correlación entre el número de habitaciones, dormitorios y el valor de la vivienda.
- Identificar patrones en los datos mediante visualizaciones.
- Extraer conclusiones sobre los factores que pueden influir en los precios de las viviendas.

*This research was supported by grant No. xxxx.

3 Análisis Univariado

Se realizó un análisis univariado para comprender la distribución de cada variable. Para ello, se generaron histogramas, diagramas de caja y gráficos de medidas de centralidad.

3.1 Housing Median Age (Edad Mediana de las Viviendas)

- **Gráfica 1:** Los datos muestran que la edad de las viviendas oscila entre 0 y 54 años. Se observa que hay casas en todo el rango de edades, pero la mayoría se encuentran entre los 15 y 36 años.
- **Gráfica 2:** Se puede afirmar que el promedio de edad de las viviendas es de 29 años, al igual que la mediana. Esto significa que el 50% de las viviendas tienen menos de 29 años y el otro 50% tienen más. En el caso de la moda, se observa que hay muchas viviendas con más de 50 años.
- **Gráfica 3:** Se observa que las edades de las viviendas se encuentran entre 0 y más de 50 años. El 50% de las viviendas están representadas en la caja del diagrama, con edades que oscilan entre 18 y 36 años aproximadamente. La media está entre 28 y 29 años, lo que confirma la concentración de viviendas en ese rango.

3.2 Total Rooms (Total de Habitaciones)

- **Gráfica 1:** La mayoría de las viviendas tienen entre 0 y 5.000 habitaciones. Sin embargo, se observa que hay algunas con hasta 25.000 habitaciones, lo que podría representar valores atípicos o errores en los datos.
- **Gráfica 2:** Se puede afirmar que el promedio de habitaciones es superior a 2.500, al igual que la mediana que está por encima de las 2.000 habitaciones. Esto significa que el 50% de las viviendas tienen menos de 2,000 habitaciones y el otro 50% tienen más. En el caso de la moda, se observa que hay muchas viviendas con cerca de 1,000 habitaciones.
- **Gráfica 3:** El 50% de las viviendas tienen entre 1.500 y 3.000 habitaciones. La media sigue siendo superior a 2500 habitaciones. Los bigotes indican que los valores típicos oscilan entre 0 y 5.100 habitaciones. Además, se observa una gran cantidad de valores atípicos, es decir, viviendas con cantidades de habitaciones muy alejadas de la media, que están fuera del rango del IQR.

3.3 Total Bedrooms (Total de Dormitorios)

- **Gráfica 1:** El número total de dormitorios oscila entre 0 y menos de 3.000. La mayoría de las viviendas tienen entre 0 y 1.000 dormitorios, con una concentración notable alrededor de los 500 dormitorios.

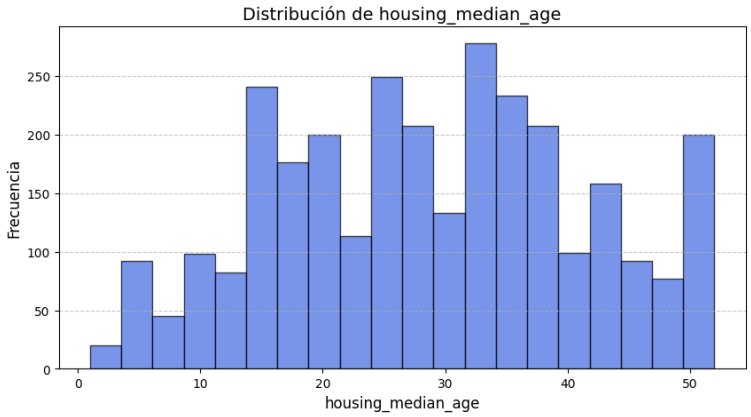

Figure 1: Distribución de la Edad Mediana de las Viviendas.

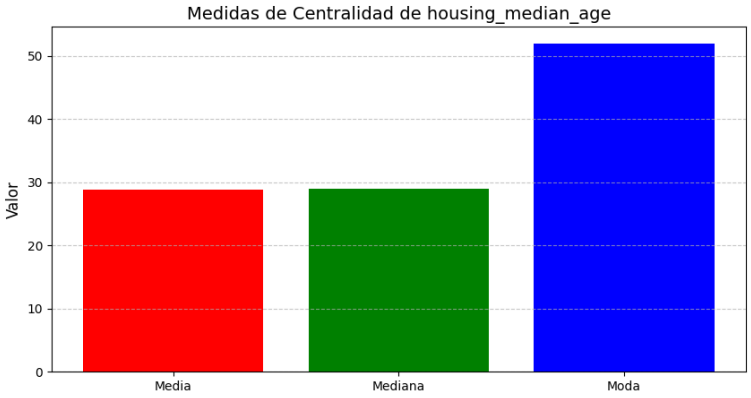

Figure 2: Medidas de Centralidad de la Edad Mediana de las Viviendas.

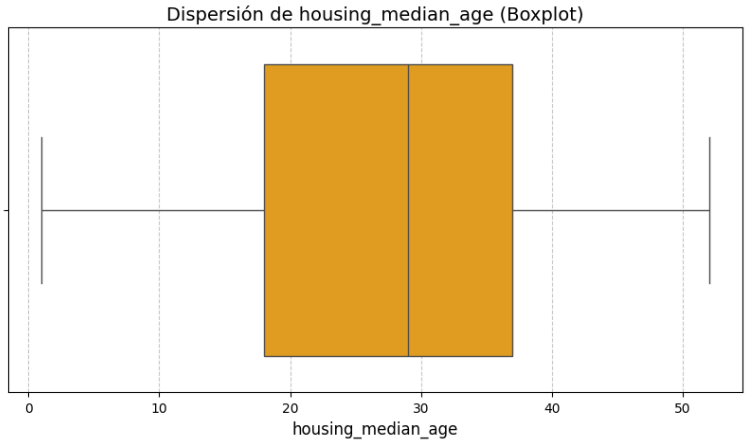

Figure 3: Diagrama de Caja de la Edad Mediana de las Viviendas.

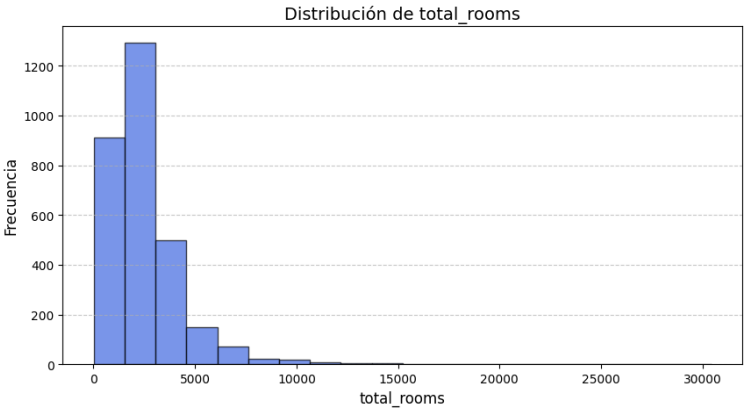

Figure 4: Distribución Total de Habitaciones.

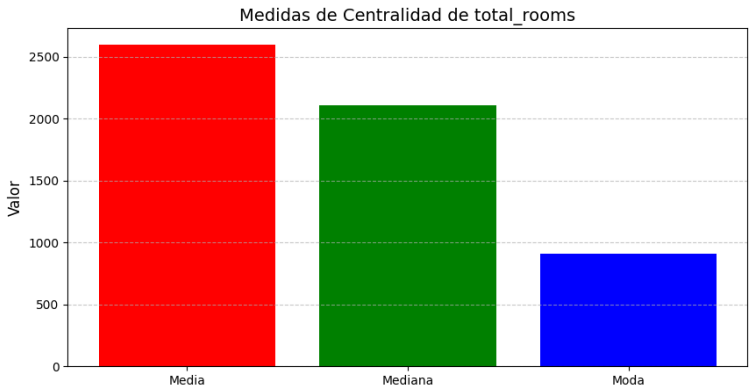

Figure 5: Medidas de Centralidad de Total de Habitaciones.

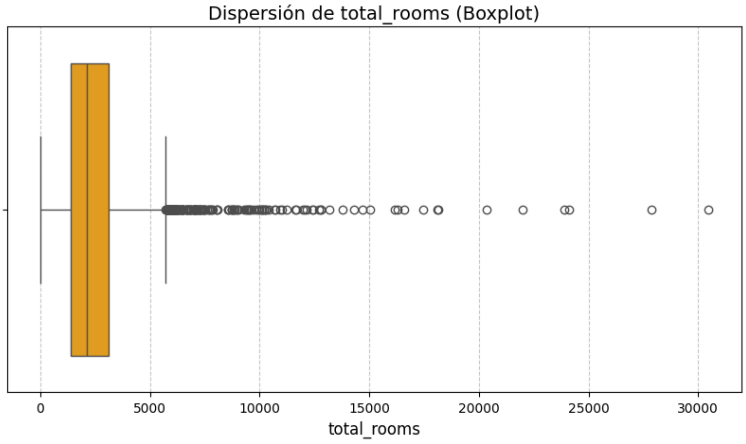

Figure 6: Diagrama de Caja de Total de Habitaciones.

- **Gráfica 2:** Se puede afirmar que el promedio de dormitorios es superior a 500, al igual que la mediana, que supera los 400 dormitorios. Esto significa que el 50

- **Gráfica 3:** El 50Existen muchos valores atípicos, con viviendas que tienen entre 1,100 y 2,000 dormitorios e incluso algunas con más de 5,000 dormitorios, lo que podría representar casos especiales o errores en los datos.

3.4 Population (Población en las zonas)

- **Gráfica 1:** La mayoría de las zonas tienen una población entre 0 y 2,000 personas. Sin embargo, hay valores que se extienden hasta 12,000 habitantes, lo que puede indicar desde ya la presencia de valores atípicos.

- **Gráfica 2:** Se puede afirmar que el promedio de población es de 1.400 personas, al igual que la mediana, que es de aproximadamente 1.200 personas. Esto significa que el 50% de las zonas tienen menos de 1.200 habitantes y el otro 50% tienen más. En el caso de la moda, se observa que muchas zonas tienen alrededor de 850 habitantes.

- **Gráfica 3:** Los valores típicos de la población oscilan entre 0 y 3.200 habitantes, el 50% de las zonas tienen una población entre 1.000 y casi 2.000 personas. Se observa una gran cantidad de valores atípicos en poblaciones superiores a 3.200 habitantes.

3.5 Households (Número de Hogares por Zona)

- **Gráfica 1:** La mayoría de las zonas tienen pocos hogares, más de 1.200 zonas cuentan con apenas 500 hogares. Y que solo algunas zonas superan los 3.000 hogares.

- **Gráfica 2:** Se puede afirmar que el promedio de hogares es menor a 500, la mediana por su parte está por encima de 400 hogares. Esto significa que el 50% de las zonas tienen menos de 400 hogares y el otro 50% tienen más. En el caso de la moda, se observa que muchas zonas tienen cerca de 380 hogares.

- **Gráfica 3:** Los valores típicos oscilan entre 0 y un poco más de 10000 hogares. Muchas zonas tienen menos de 10000 hogares, pero hay algunas con más de 2.000 o incluso 50000 hogares, lo que las convierte en valores atípicos.

3.6 Median Income (Ingreso Mediano por Zona)

- **Gráfica 1:** La mayoría de las zonas tienen ingresos medianos entre 2 y 5, pero existen algunas con ingresos significativamente más altos.

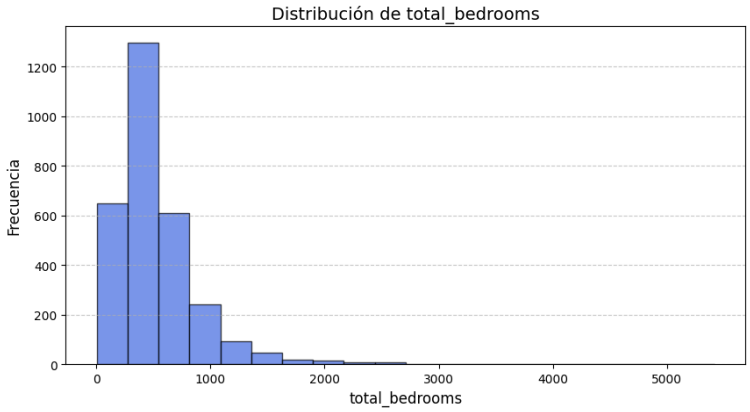

Figure 7: Distribución Total de Dormitorios.

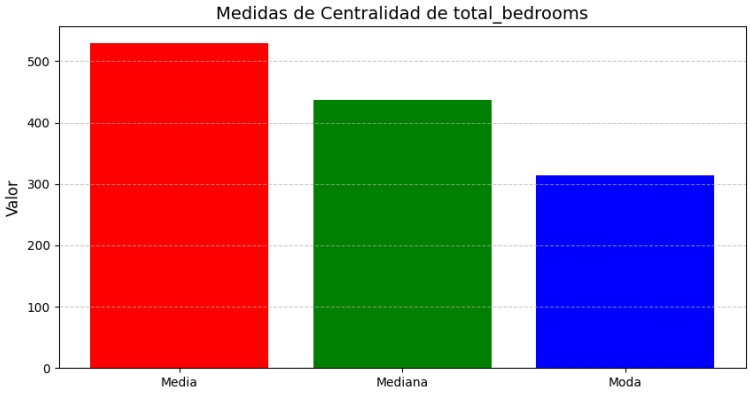

Figure 8: Medidas de Centralidad de Total de Dormitorios.

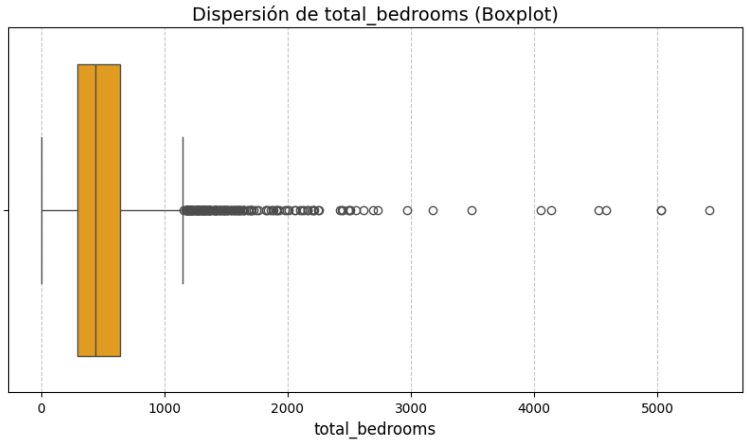

Figure 9: Diagrama de Caja de Total de Dormitorios.

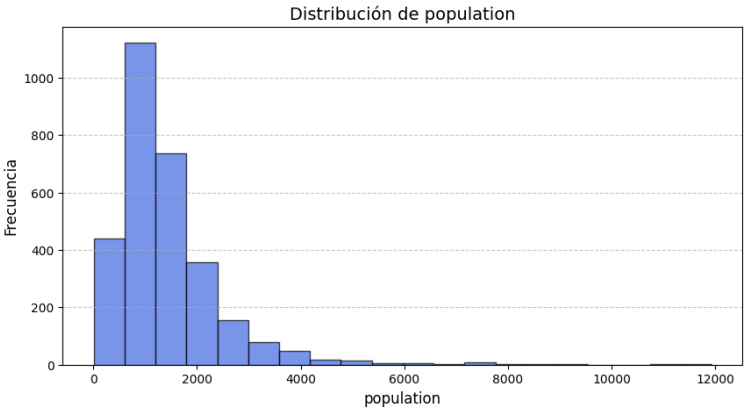

Figure 10: Distribución de la Población en las zonas.

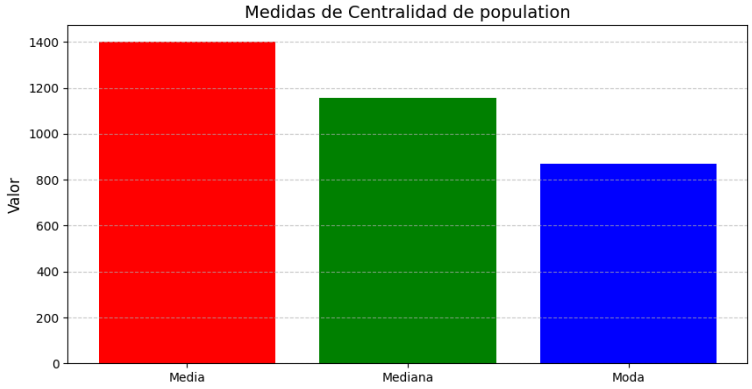

Figure 11: Medidas de Centralidad de la Población en las zonas.

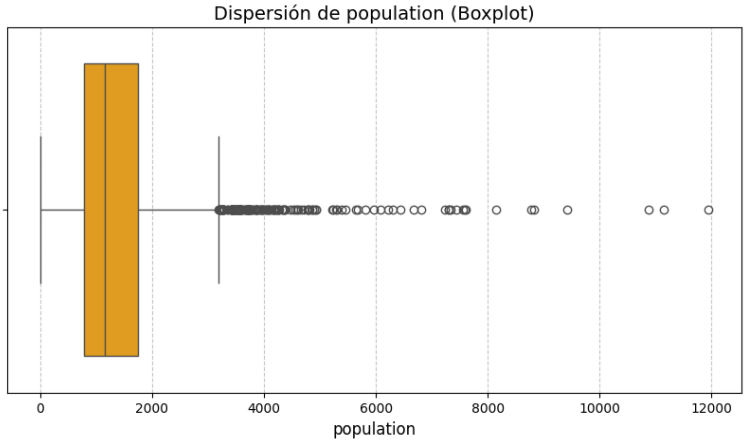

Figure 12: Diagrama de Caja de la Población en las zonas.

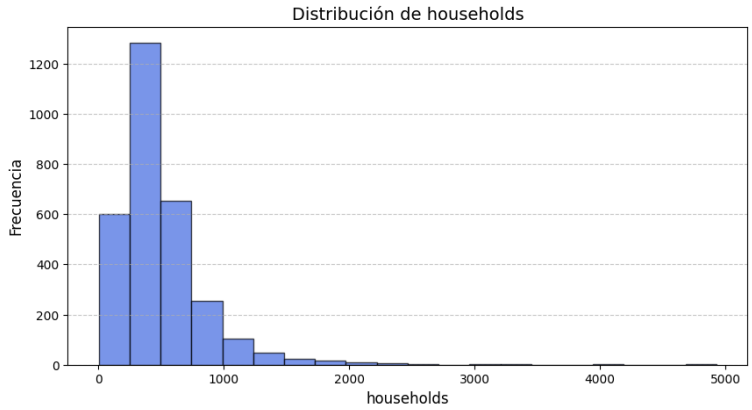

Figure 13: Distribución de Número de Hogares por Zona.

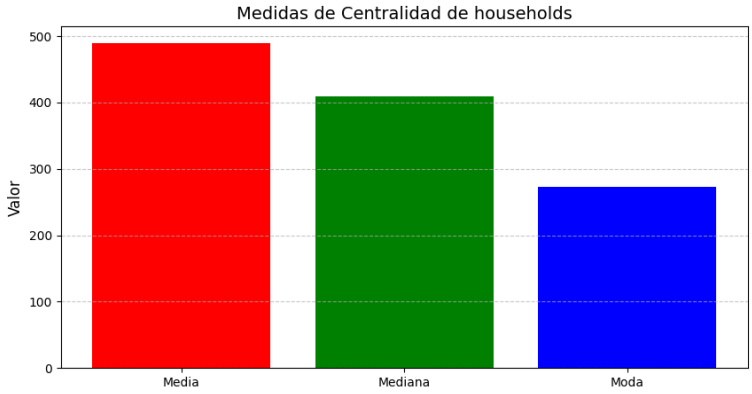

Figure 14: Medidas de Centralidad de Número de Hogares por Zona.

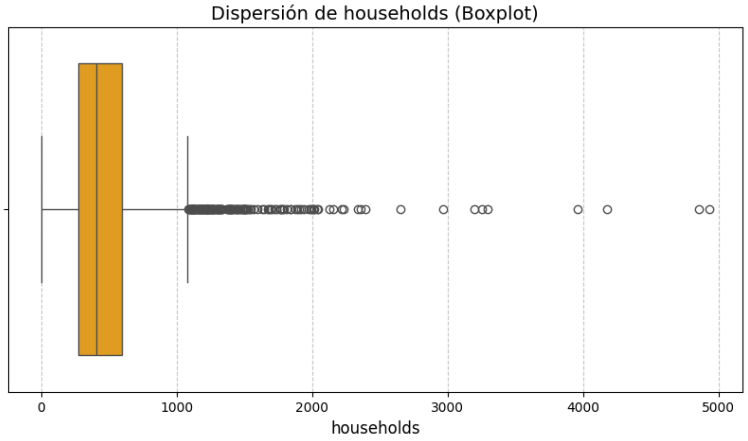

Figure 15: Diagrama de Caja de Número de Hogares por Zona.

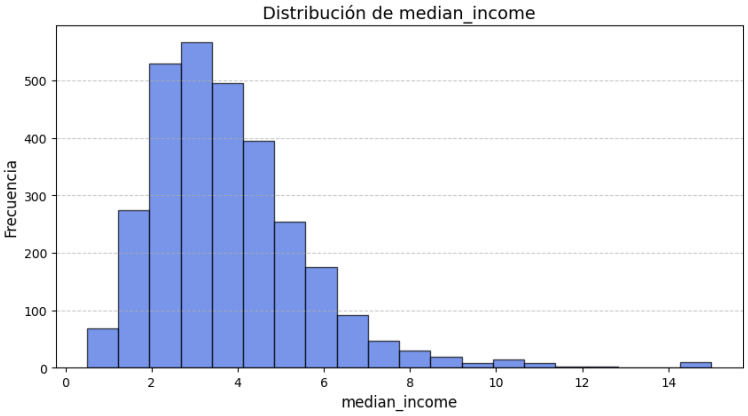

Figure 16: Distribución de Ingreso Mediano por Zona.

- **Gráfica 2:** Se puede afirmar que el promedio de ingreso mediano es aproximadamente 4, al igual que la mediana. Esto significa que el 50
- **Gráfica 3:** : Los valores típicos oscilan entre 0 y menos de 8, el 50

3.7 Median House Value (Valor Mediano de las Viviendas)

- **Gráfica 1:** La mayoría de las zonas tienen ingresos medianos entre 2 y 5, pero existen algunas con ingresos significativamente más altos.
- **Gráfica 2:** La mayoría de las viviendas tienen valores entre 100.000 y 200.000 dólares. A medida que aumenta el valor de las viviendas, la frecuencia disminuye, hasta llegar a 5000000, donde hay un aumento significativo en este valor.
- **Gráfica 3:** : : La mayoría de las viviendas tienen valores entre 1000.000 y 300.0000 dólares, que viviendas con un valor de 500.000 dólares se consideran valores atípicos según los valores y el calco del IQR.

4 Análisis Multivariado

4.1 Análisis de la Matriz de Correlación

total_rooms y total_bedrooms están altamente correlacionados (0.94), lo que nos dice que ambas proporcionan información similar, que median_house_value tiene una relación muy débil con ambas variables, lo que sugiere que el número de habitaciones o dormitorios no es un factor clave para determinar el precio de la vivienda.

4.2 Relación entre Total de Habitaciones y Valor de la Vivienda

El gráfico muestra que no hay una relación entre el número de habitaciones y el valor de la vivienda, ya que los datos se encuentran dispersos y no hay una tendencia definida. La mayoría de las viviendas tienen menos de 5.000 habitaciones, pero incluso entre ellas, el valor de las casas varía ampliamente.

4.3 Relación entre Total de Habitaciones y Total de Dormitorios

Este gráfico nos muestra una fuerte correlación positiva entre el número total de habitaciones y el número total de dormitorios, con una tendencia lineal. A medida que aumenta la cantidad de habitaciones, también lo hace la cantidad de dormitorios, lo que tiene sentido ya que una mayor cantidad de habitaciones en una vivienda generalmente implica más dormitorios.

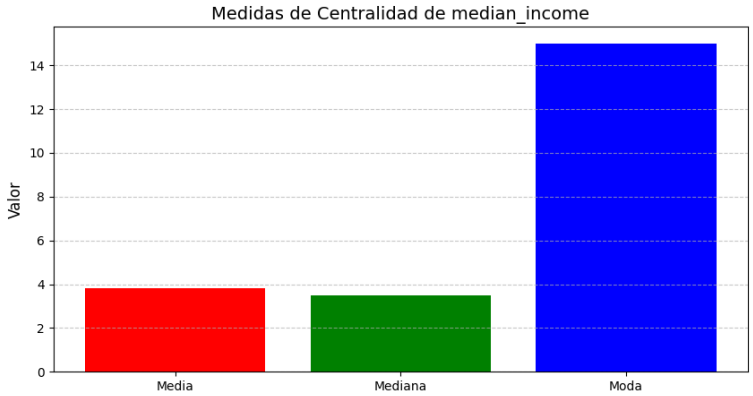

Figure 17: Medidas de Centralidad de Ingreso Mediano por Zona.

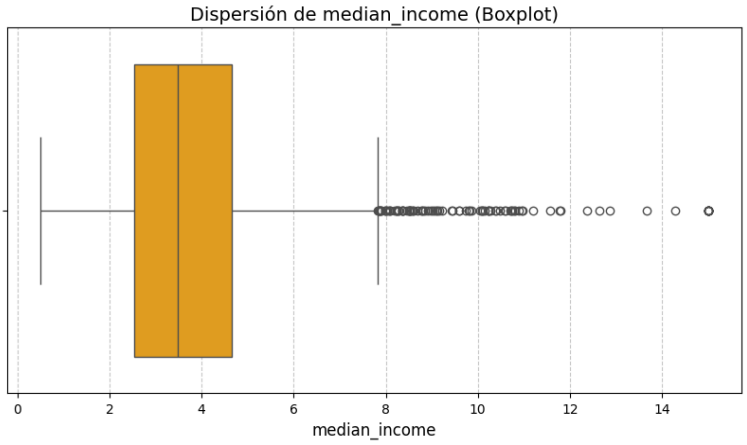

Figure 18: Diagrama de Caja de Ingreso Mediano por Zona.

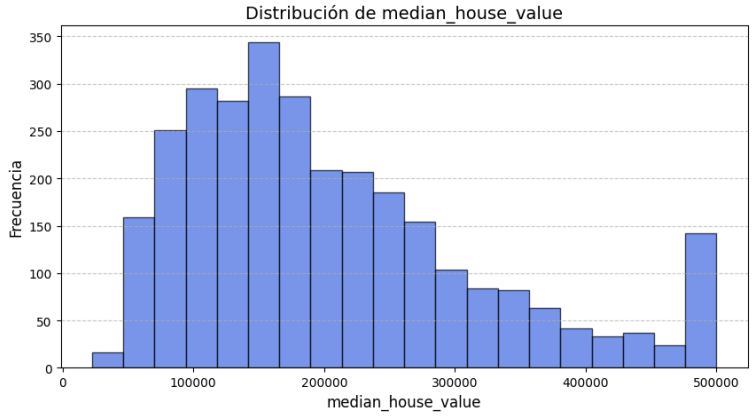

Figure 19: Distribución de Valor Mediano de las Viviendas.

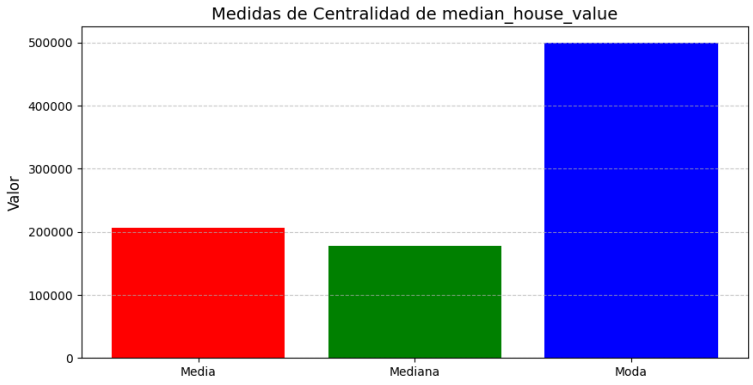

Figure 20: Medidas de Centralidad de Valor Mediano de las Viviendas.

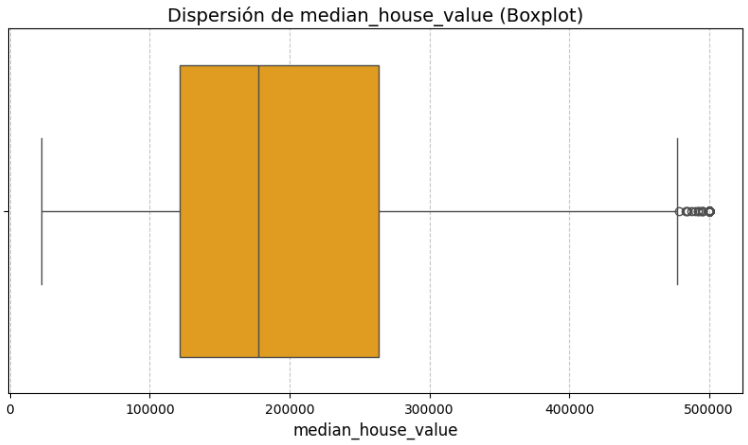

Figure 21: Diagrama de Caja de Valor Mediano de las Viviendas.

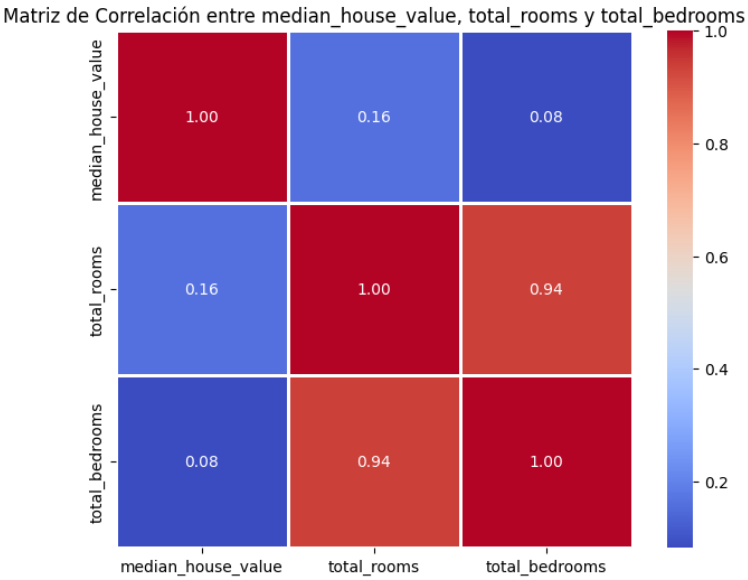

Figure 22: Matriz de Correlación entre total_rooms, total_bedrooms y median_house_value.

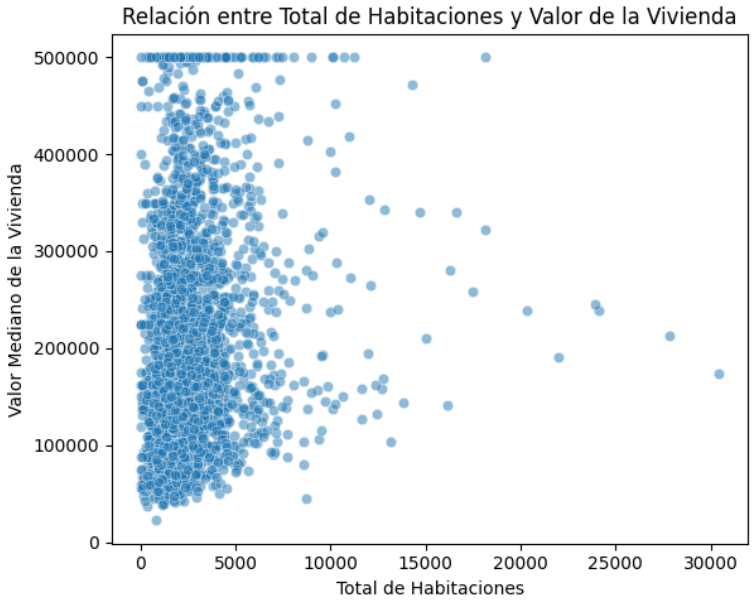

Figure 23: Relación entre Total de Habitaciones y Valor de la Vivienda.

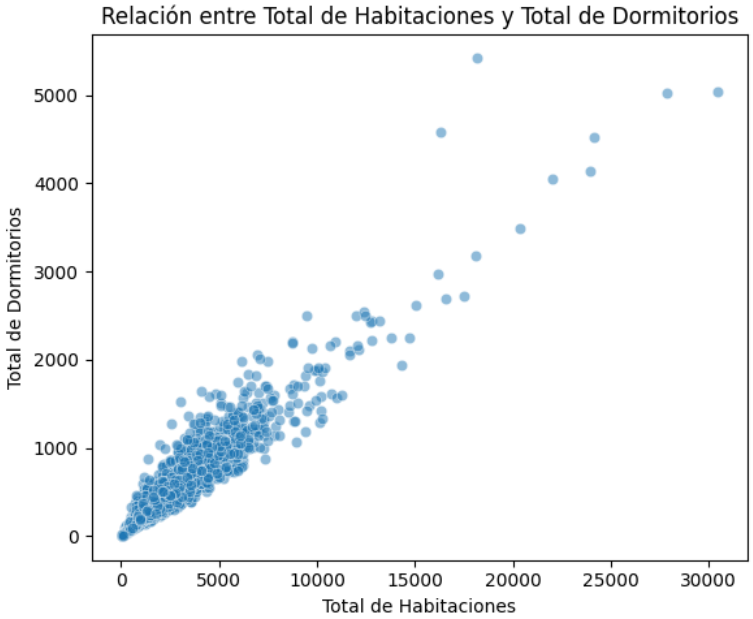

Figure 24: Relación entre Total de Habitaciones y Total de Dormitorio.

5 Conclusiones

El análisis realizado muestra que ni el número de habitaciones ni el número de dormitorios tienen una fuerte influencia en el valor de la vivienda en California. Sin embargo, estas dos variables están altamente correlacionadas entre sí, lo que indica que una de ellas podría ser eliminada sin perder información significativa en un análisis predictivo. Otras variables, como la ubicación o el ingreso mediano de la zona, podrían ser factores más determinantes en el valor de las viviendas. Estos resultados sugieren que un estudio más profundo, incluyendo otras variables, podría proporcionar una mejor comprensión de los factores que afectan los precios de las viviendas.

6 Referencias

- `https://www.youtube.com/watch?v=7GzcFTWPAds` - Análisis multivariable, ¿de qué se trata? Básico y simplificado.
- `https://www.youtube.com/watch?v=RzPzqvkTZRU&t=387s` - Análisis univariable, ¿de qué se trata? Básico y simplificado.