

Análisis Estadístico de Datos de Vivienda en California

Juan Orozco, Diego Fernandez, Nayelis Jaimes

March 3, 2025

Abstract

Este reporte presenta un análisis estadístico del conjunto de datos de vivienda en California. Se realizan análisis univariados y bivariados para extraer información relevante sobre la distribución y relación entre variables. Se incluyen histogramas, gráficos de dispersión y medidas estadísticas clave.

0.1 Objetivos de los Análisis

Nuestro objetivo principal de nuestro trabajo es investigar de manera practica y fácil, la manera en que se comportan las variables en nuestro conjunto de datos de las viviendas en california. Buscamos la manera de comprender como distribuir las características principales (como la edad de las viviendas, el numero de habitaciones y el valor de las propiedades) al momento de identificar patrones y casos atípicos que puedan influir en el mercado, también se tuvo en cuenta el analizar la manera de como se concentraban los datos y su dispersión para ver tendencias generales y explorar, las relaciones entre variables de suma importancia como las que serian :(cuantas habitaciones podría aumentar de manera proporcionar el valor de la casa, para ser preciso este análisis nos permitirá obtener una visión clara y fundamental del comportamiento sobre las zonas inmobiliarias dándonos nuevas ideas útiles para la toma de decisiones en situaciones de de la realidad dando un enfoque practico para la tarea universitaria

0.2 Gráficos

0.2.1 Histogramas de Variables

A continuación, se presentan histogramas de las variables analizadas.

0.2.2 Gráficos de Dispersión

A continuación, se presentan gráficos de dispersión para visualizar la relación entre variables.

0.2.3 Gráficos de Dispersión

A continuación, se presentan gráficos de dispersión para visualizar la relación entre variables.

0.3 Análisis de los Resultados

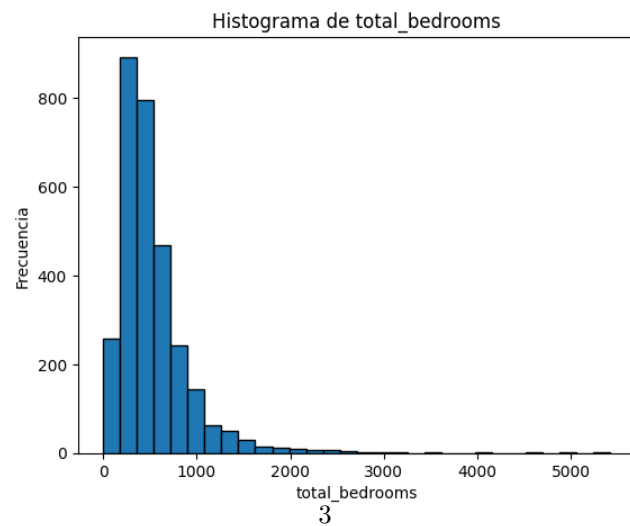
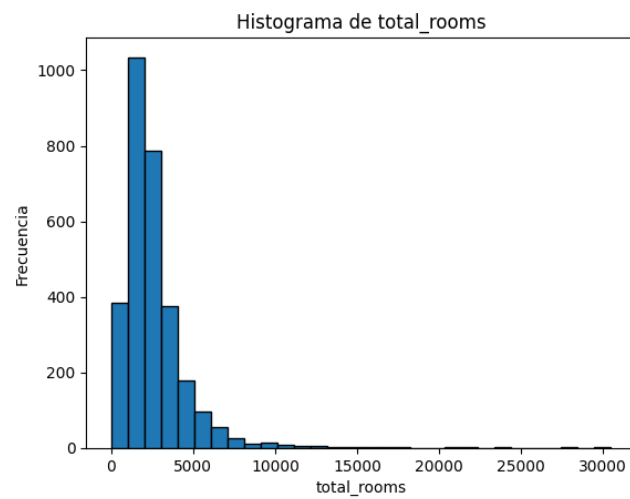
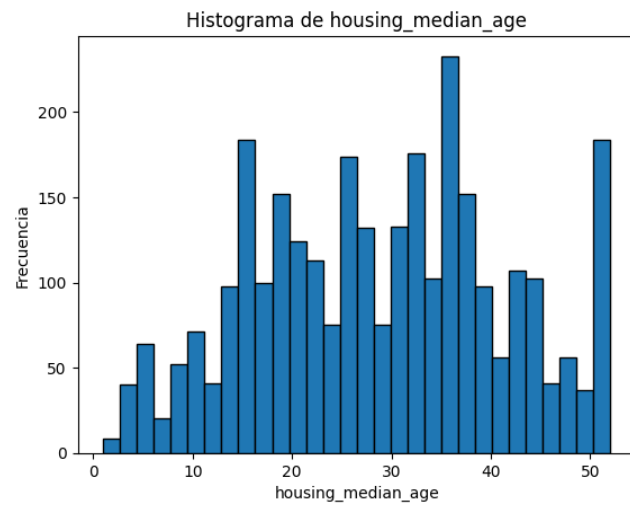
Cuando observamos los pictogramas, se hace evidente cómo se agrupan los datos en diferentes rangos. Por ejemplo, al analizar la antigüedad de las viviendas, notamos que la mayoría de ellas se encuentra en un rango medio. Esto sugiere que muchas casas tienen una edad similar. En cuanto al número de habitaciones, vemos que la mayoría tiene un número moderado, aunque hay algunas que destacan porque tienen muchos más cuartos. Esto nos da una idea clara de dónde se concentra la mayoría de los datos y también de aquellos casos especiales que se alejan de la norma.

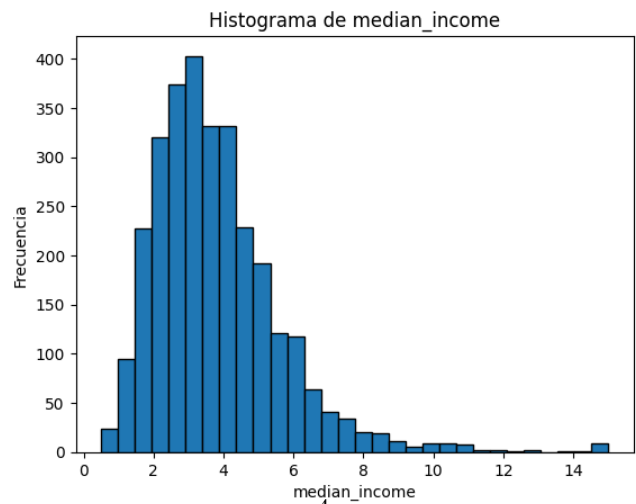
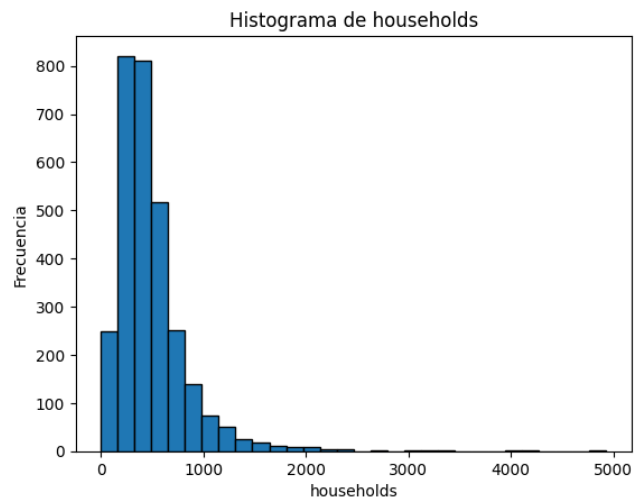
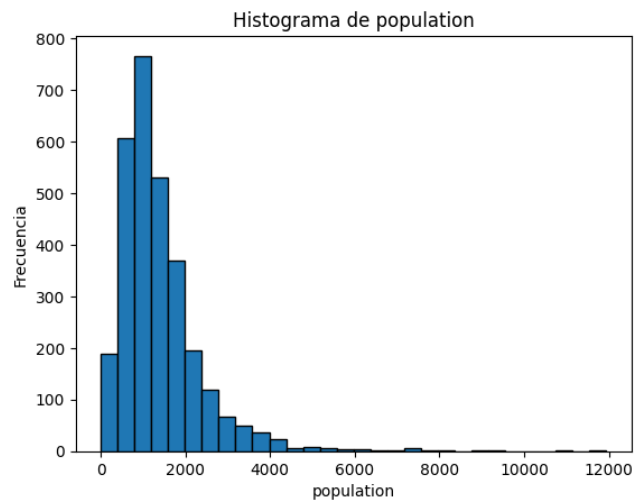
Los boxplots son herramientas muy útiles para comprender la centralidad de los datos. Nos muestran de manera sencilla dónde está el centro y cómo se

distribuyen los valores a su alrededor. Imagina una “caja” que agrupa la mayor parte de los datos, con una línea en el medio que marca el punto central. En el caso de la antigüedad de las viviendas, esta caja se ve compacta, lo que indica que muchas casas son similares en edad. Por otro lado, cuando observamos el total de habitaciones, notamos algunos puntos que sobresalen de la caja, lo cual nos revela que hay casas con características muy distintas a las demás. Es una forma práctica y visual de entender dónde se concentra la mayoría y cuáles son esos casos excepcionales.

Al analizar las medidas de dispersión a través de gráficos que muestran la media y la mediana, podemos ver no solo dónde está el centro de los datos, sino también qué tan dispersos están. En estos gráficos, se dibujan dos líneas: una representa la media y otra la mediana, sobre una curva que muestra cómo se distribuyen los datos. Si ambas líneas están muy cerca una de la otra, significa que los datos tienden a agruparse en torno a un valor similar. Si por el contrario, la media se desplaza hacia un lado, eso indica que hay algunos valores extremos que están influyendo en el promedio. Esta visualización nos ayuda a comprender mejor cómo se comportan los datos y a identificar aquellos casos que se desvían del patrón general.

Por último, el análisis bivariado nos permite explorar las relaciones entre dos variables para ver cómo interactúan entre sí. Por ejemplo, al comparar el número de habitaciones con el valor de las casas, generalmente observamos que a medida que aumenta el número de habitaciones, también lo hace el precio; esto sugiere una relación directa entre ambas variables. Sin embargo, al comparar la antigüedad de las viviendas con su precio, notamos una mayor dispersión en los datos, lo que indica que la antigüedad por sí sola no explica completamente las variaciones en el valor. Estos gráficos y cálculos nos ayudan a entender mejor qué factores pueden estar relacionados y cuáles son los aspectos más relevantes en el comportamiento del mercado inmobiliario.





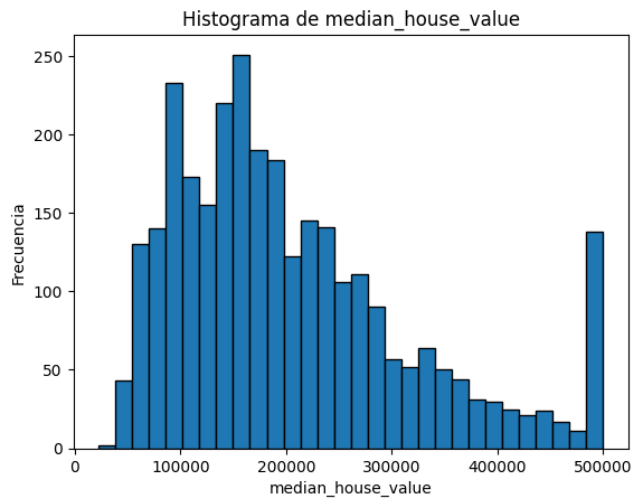


Diagrama de Caja de housing_median_age

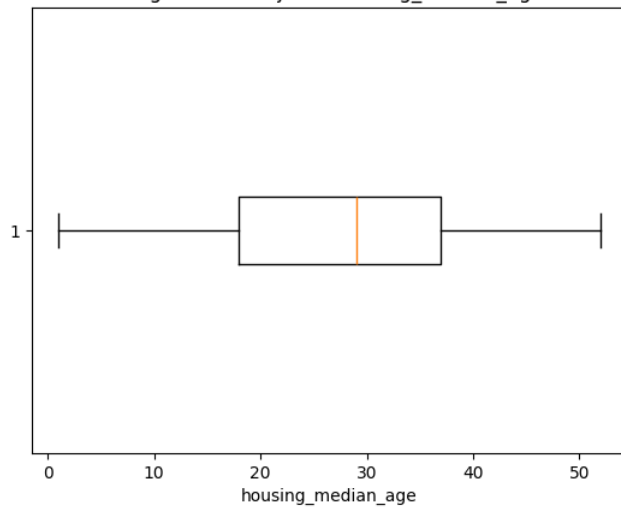


Diagrama de Caja de total_rooms

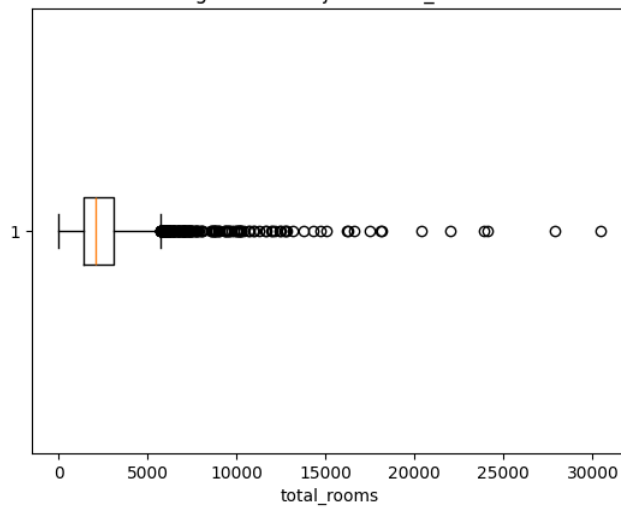


Diagrama de Caja de total_bedrooms

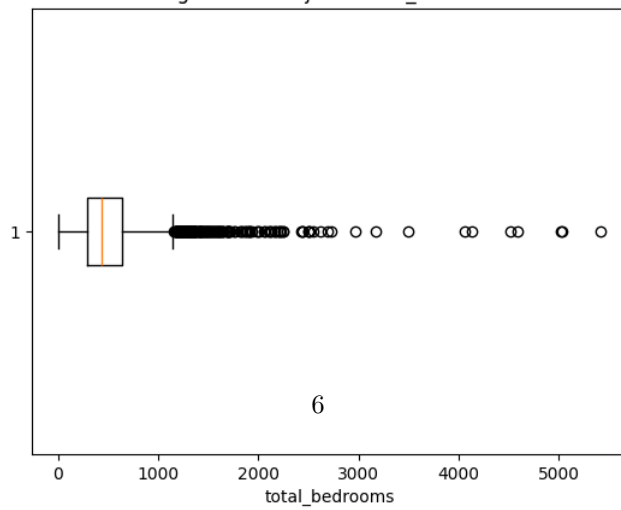


Diagrama de Caja de population

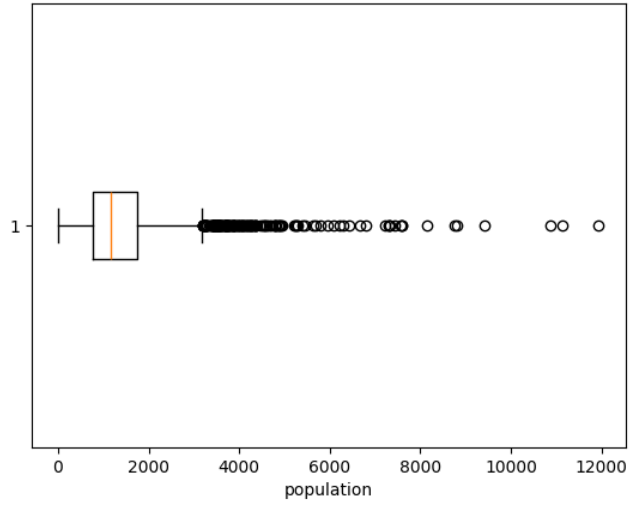


Diagrama de Caja de households

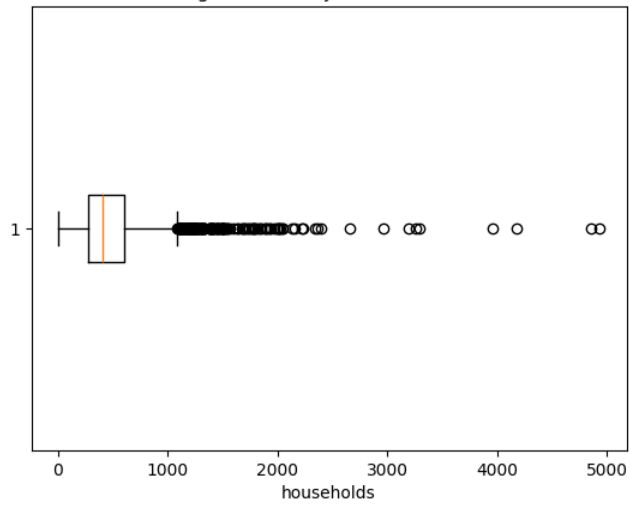


Diagrama de Caja de median_income

