Proyecto final

Miguel, Alma, Yesenia Luna Moreno

Introducción

En el mercado inmobiliario actual se generan enormes cantidades de información sobre los precios, las características físicas y la ubicación de los inmuebles. Analizar estos datos es fundamental para detectar tendencias, clasificar propiedades según sus atributos y facilitar la toma de decisiones en procesos de renta o inversión.

Para este proyecto se trabaja con el conjunto de datos Apartment for Rent Classified, disponible en el repositorio de UCI Machine Learning. Dicho dataset reúne miles de registros de apartamentos en renta, con variables como el precio, el tamaño en metros cuadrados, el número de habitaciones, baños y diferentes amenidades.

El objetivo principal es realizar un Análisis Exploratorio de Datos (EDA) que permita comprender la estructura del dataset y, después, aplicar técnicas de clustering para agrupar los apartamentos en categorías con características similares. Para esto se utilizan métodos como K-means y clustering jerárquico, apoyados en herramientas como el método del codo y el coeficiente silhouette, que ayudan a definir cuántos grupos son los más adecuados.

Con este análisis se espera obtener una visión más clara del mercado de renta, identificar distintos segmentos de apartamentos y señalar diferencias importantes entre ellos. Los resultados pueden servir como apoyo para entender cómo se distribuyen las propiedades y qué factores influyen en la formación de grupos, lo cual resulta útil para empresas inmobiliarias, arrendadores y personas interesadas en encontrar opciones de vivienda.

**Objetivo general**

Evaluar el conocimiento adquirido en el curso mediante el análisis de un conjunto de datos del mercado inmobiliario, aplicando técnicas de exploración y agrupamiento que permitan identificar patrones y segmentos de apartamentos en renta.

**Objetivos específicos**

1. Realizar un análisis exploratorio de datos (EDA) para comprender la estructura y las variables del dataset Apartment for Rent Classified.
2. Aplicar técnicas de clustering, como K-means y clustering jerárquico, con el fin de agrupar los apartamentos según sus características.
3. Determinar el número óptimo de grupos utilizando criterios como el método del codo y el coeficiente silhouette.
4. Interpretar los resultados obtenidos para identificar diferencias y similitudes entre los segmentos de apartamentos.
5. Presentar conclusiones que muestren cómo este tipo de análisis puede apoyar la toma de decisiones en el mercado de renta de inmuebles.

Antecedentes

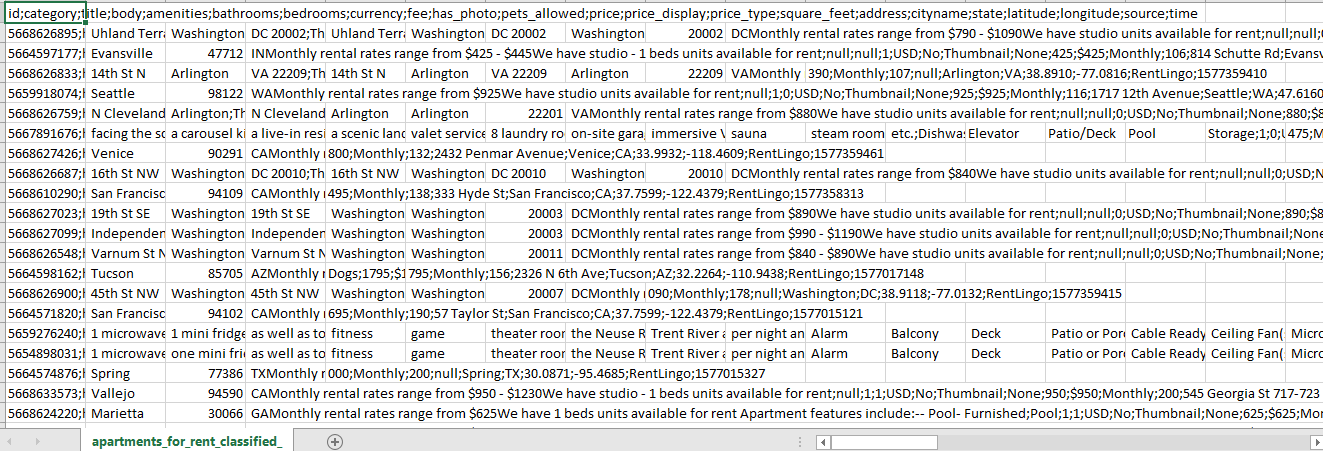
Data ser designado

Nombre: apartments\_for\_rent\_classified\_10K

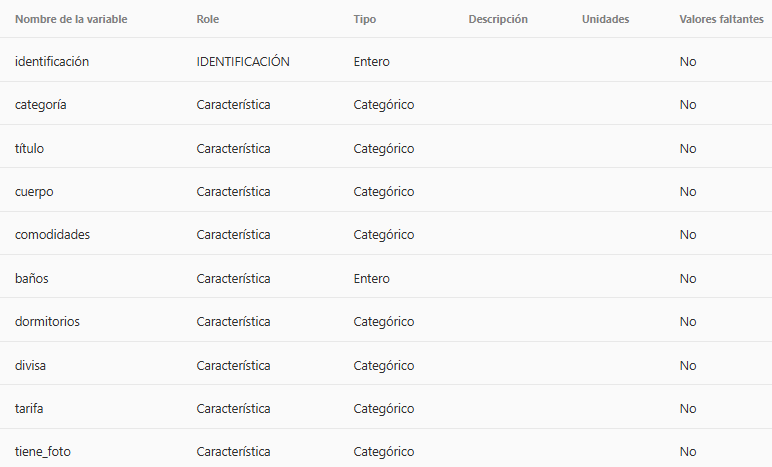
Fuente: <https://archive.ics.uci.edu/dataset/555/apartment+for+rent+classified>

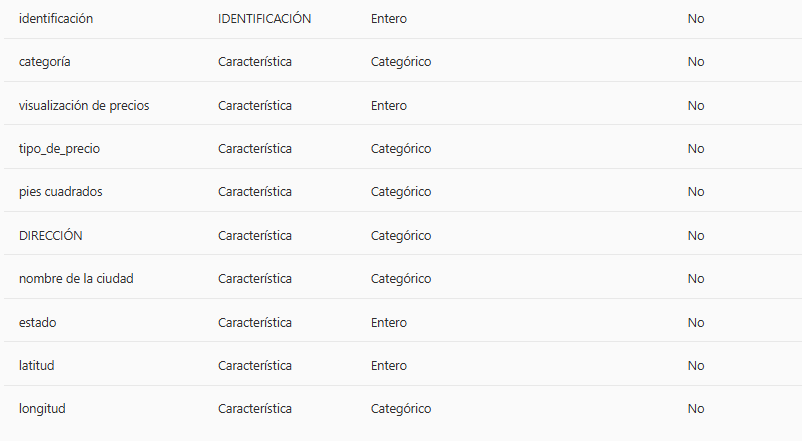
Contenido: Conjunto de datos clasificados de apartamentos en alquiler en EE.UU. contiene 10 000 filas y 22 columnas.

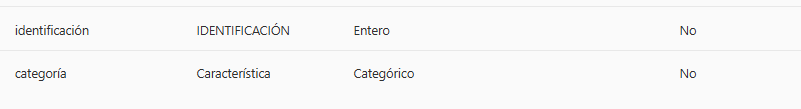
Nota: Los datos se han depurado de forma que las columnas “precio” y “pies” nunca estén vacías, No tiene valores faltantes.



Descripción de tabla de variables:



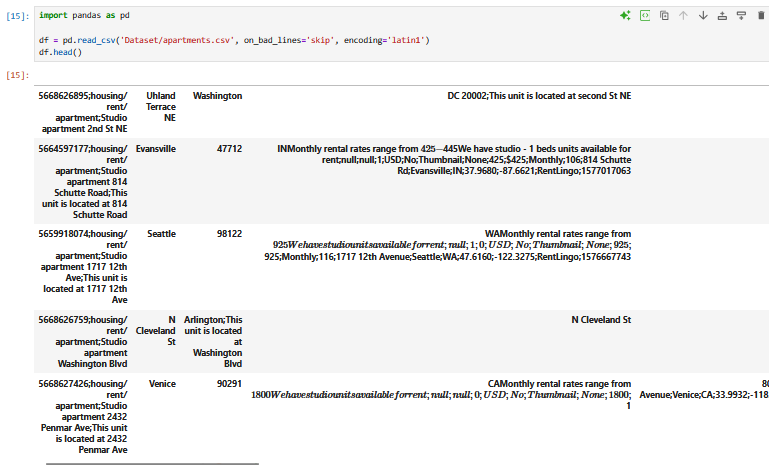




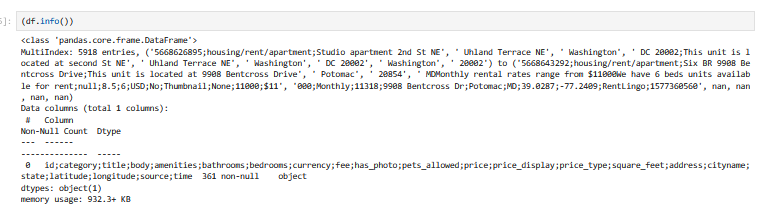
Información complementaria

id = identificador único del apartamento, category = categoría del clasificado, title = texto del título del apartamento, body = texto del cuerpo del apartamento, Amenities = como aire acondicionado, baloncesto, cable, gimnasio, acceso a Internet, piscina, refrigerador, etc. Bathrooms = número de baños, Bedrooms = número de dormitorios, Currency = precio actual, fee = tarifa, has\_photo = foto del apartamento, pets\_allowed = qué mascotas están permitidas dogs/cats etc. price = precio de alquiler del apartamento, price\_display = precio convertido a pantalla para el lector, price\_type = precio en USD, square\_feet = tamaño del apartamento, Address = dónde está ubicado el apartamento, Cityname = dónde está ubicado el apartamento, State = dónde está ubicado el apartamento, Latitude = dónde está ubicado el apartamento, Longitude = dónde está ubicado el apartamento, Source = origen del clasificado, Time = cuándo se creó el clasificado

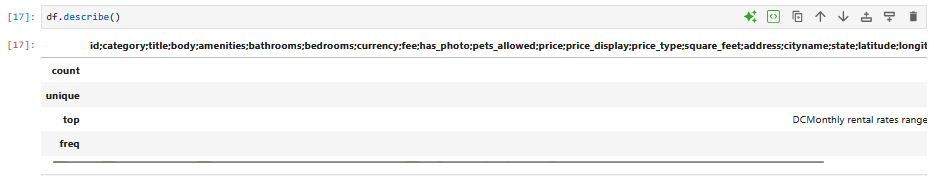
Exploración de datos



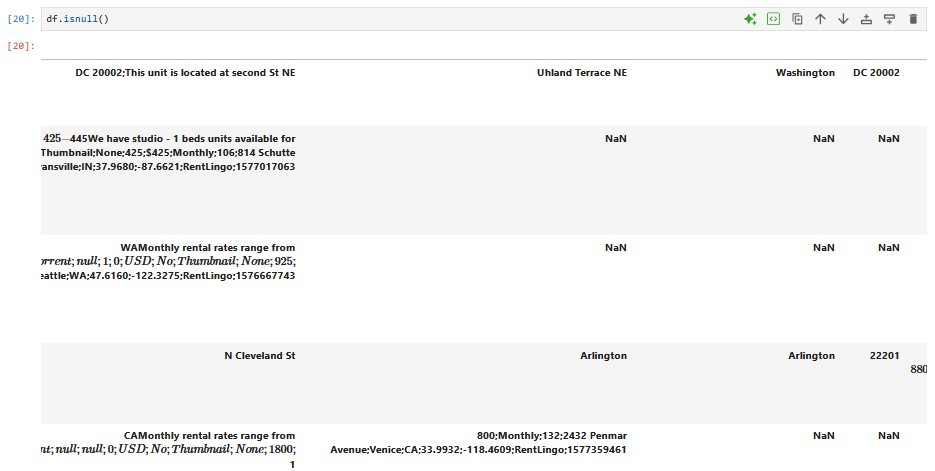
Tipos de datos



Resumen estadístico

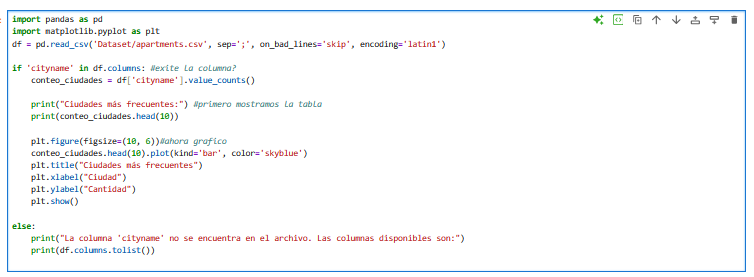


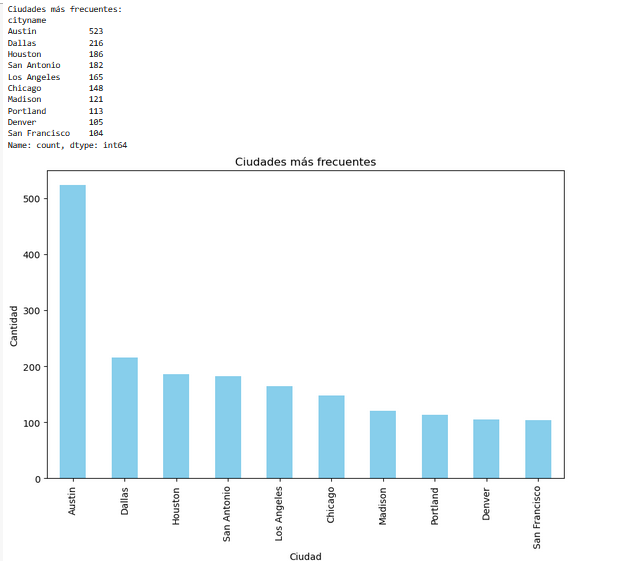
Revisión de valores nulos



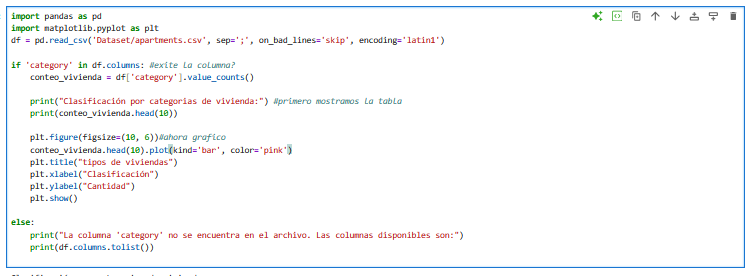
Obtenga al menos 5 insights de sus resultados. Se deben incluir gráficos interesantes que muestren el comportamiento de los datos y cómo se relacionan, así como acompañar los gráficos con una explicación de lo que se observa en el gráfico.

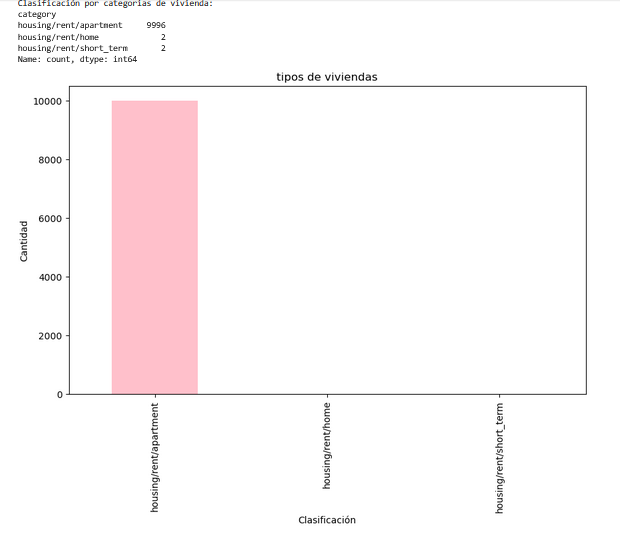
1. Ciudades más frecuentes



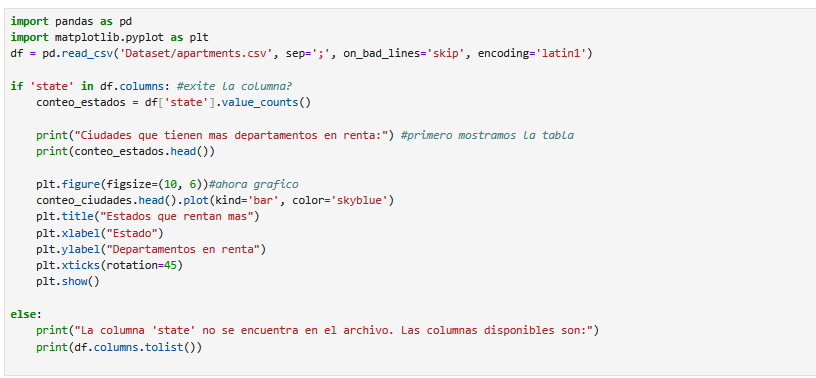


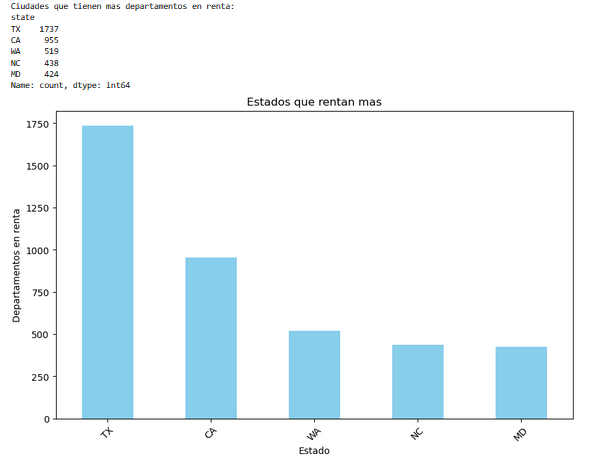
1. Clasificación por tipo de vivienda o categoría de vivienda





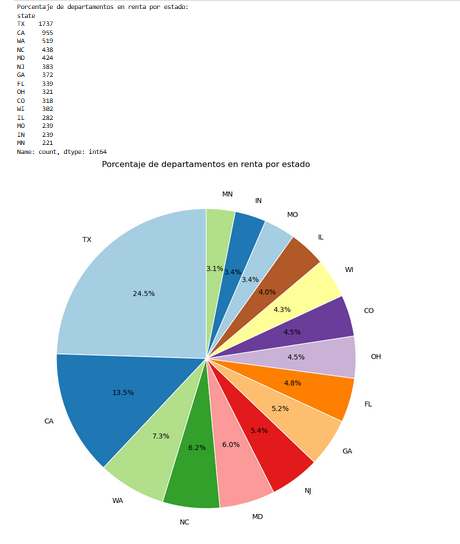
1. Ciudades con mayor cantidad de departamentos en renta





1. Porcentaje de departamentos que se rentan por estado





1. Distribución de precios: Visualiza cómo se distribuyen los precios de los apartamentos en el conjunto de datos.
2. Distribución de habitaciones: Muestra la frecuencia de apartamentos con diferentes cantidades de habitaciones.
3. Relación entre precio y número de habitaciones: Examina si existe alguna relación entre el precio de los apartamentos y el número de habitaciones que tienen.
4. Distribución de precios por estado: Muestra cómo varían los precios de los apartamentos según el estado en el que se encuentren.
5. Cada gráfico proporciona una visión diferente sobre los datos, ayudando a entender patrones y relaciones dentro del conjunto de datos de los apartamentos en renta.

