

Apartments for rent classified Analysis

Source of data: The data was downloaded from Machine Learning Repository website and only one of the files of 100K rows will be analysed and includes geographic variables.

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

Data collection: The dataset is from 2019 and it was saved as it was created. It's licensed under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0) license and allows for sharing and adaptation for any purpose.

Data content: It contains 22 columns more than 100,000 rows. The variables are amenities, number of bathrooms, number of bedrooms, currency, photo of the apartment, what pets are allowed, price, size of the apartment, where the apartment is located.

Data Limitations:

- There is little information about how the data was collected.
- Dataset contains many null and blanks values which will make the data not relevant since the percentage is too high.
- Because the data is not collected for a few years in a row it doesn't allow comparisons and predictions.
- Information about whether the apartment was rented would have been useful to get more insights and make price-based renting decisions.
- It's not specified anywhere if the column "Fee" refers to the realtor fees or condo fees.

Data Ethics: there are no concerns because the data doesn't contain any personal information. Therefore, no PLA security is required.

Data Profile

Initially, the dataframe had 100,000 rows and 21 columns.

After a fast quality check in excel, the dataset remained with 99,117 rows and 19 columns.

84 duplicates were removed in excel.

Column "Bathroom" had 87 entries with \$1 – removed- there was no other info in the other columns

27 entries with \$2 – deleted – no other info under the other columns

6 with \$3 – removed – no other info in the other columns

1 with \$4 – removed – no other info in the other columns

1 with \$5 – removed – no other info in the other columns

1 with \$15 – removed – no other info in the other columns

Moved 420 USD under the column Bedroom to Currency and replaced USD with null under Bedrooms

“Has Photos” column had two codes “No” and “None” and none was replaced with no for consistency

In the “Has Pets” column there were 4013 null and blanks entries that were replaced with 99 979799 blanks under State – no latitude and longitude either or city and address – they have been replaced with 999

There are many null values for columns such as “Amenities”, “Bathrooms,” “Bedrooms,” “Fee”, “Has photo”. These have been investigated further in Jupyter Notes and kept as is in the end.

<i>Variable</i>	<i>Description</i>	<i>Quant/Qual</i>	<i>Structured/ Unstructured</i>	<i>Time-variant/invariant</i>
Id	Unique identifier of apartment	Qualitative	Unstructured	Invariant
City	City where the rental is located	Qualitative	Structured	Invariant
Postal Code	Postal Code where the rental is located	Qualitative	Unstructured	Invariant
Amenities	AC, basketball, cable, gym, internet access, pool, refrigerator	Qualitative	Structured	Variant
Bathrooms	Number of bathrooms	Quantitative	Structured	Invariant
Bedrooms	Number of bedrooms	Quantitative	Structured	Invariant
Currency	Price in currency	Qualitative	Structured	Invariant
Fee	Fee	Quantitative	Unstructured	Variant
Has_Photo	Photo_of_apartment	Qualitative	Structured	Invariant
Pets_Allowed	What pets are allowed dogs/cats	Qualitative	Unstructured	Invariant
Price	Rental price of apartment	Qualitative	Unstructured	Invariant
Price_Display	Price converted into display for reader	Qualitative	Unstructured	Invariant
Price_Type	Price in USD	Qualitative	Unstructured	Invariant
Square_Feet	Size of the apartment	Quantitative	Unstructured	Invariant
Address	Where the apartment is located	Qualitative	Unstructured	Invariant
Cityname	Where the apartment is located	Qualitative	Structured	Invariant
State	Where the apartment is located	Qualitative	Structured	Invariant
Latitude	Where the apartment is located	Quantitative	Unstructured	Invariant
Longitude	Where the apartment is located	Quantitative	Unstructured	Invariant
Source	Origin of classified	Qualitative	Structured	Invariant
Time	When classified is created	Quantitative	Unstructured	Invariant

Questions to explore:

1. How the rental price varies across the states?
2. Where are the cheapest rentals?
3. What type of rentals are the most popular?
4. What is the distribution of property sizes?