

REAL ESTATE VALUATION CONSULTING

Properati Real-State

- Pooja Bera
- Nataly Dirlay Martínez Romero
- Prajakta Digambar Sirame
- Carolina Velásquez Velásquez

PROJECT OBJECTIVE

WHAT IS PROPERATI?

It is a web and mobile platform for properties that was born to change the way of selling and leasing real estate in Latin America, and also seeks to provide information to help in decision-making.

for more information: https://www.properati.com.ar/acerca_de

WHAT DO THEY NEED?

Create a valuer based on the information of each property, in a way that facilitates the process for the appraisers and making it less subjective.



QUESTIONS TO BE ANSWERED

FINDING VALUE

- What is the size of the dataset?
- How many rows and columns does it has?
- How many missing values are in each column?
- How is the distribution of each variable?
- How are the variables related to each other?
- How can be the data be represented by graphs?
- How are the numerical variables correlated? What type of graph will be suitable to present this information? *What will be the best predictors of the variable of interest?

DATASET INFORMATION AND TYPE

Date	Numerical (float)	Object/Text
start_date		
end_date	lat lon rooms bedrooms bathrooms surface_total surface_covered price	1 2 3 currency title description property_type operation_type operation_type
created_on		

PROJECT

JUPITER NOTEBOOK

Methodology



DATA CLEANING

PROPERATI DATABASE

- Lat, Lon: We decided to drop the columns latitude and longitude, since those were columns related to geographic location, and we are not going to displays maps in this project.
- Bathrooms, surface_covered and surface_total: We decided to replace the null values by the mean of the data according to the property_type

DATA FRAME MANIPULATION

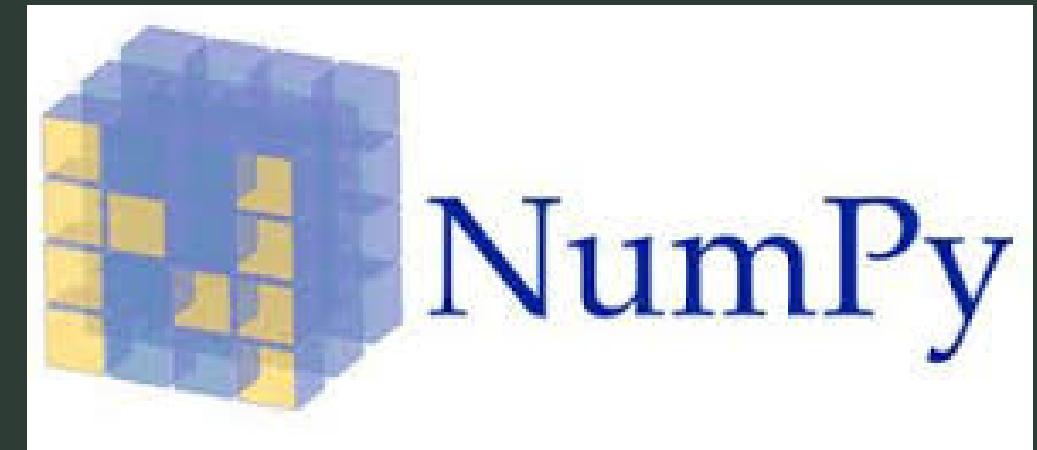
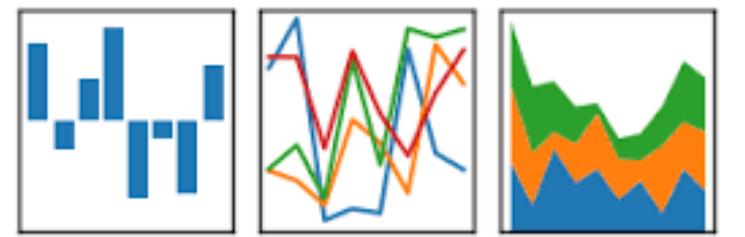
Concerning data frame and data analysis, this project show in general terms how to extract specific values from the database and perform mathematical operations with them, whether to filter, delimit, change, analyze, and manipulate information clearly and easily. for the execution of this point, the Pandas and NumPy library were imported

Below are the functions and codes applied:

DATA ANALYSIS

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



- Filter according to the city and the property's neighborhood, calculate the average of the total surface.
- How to display only the columns related to the location of each property.
- Based on a specific budget, classify the place as “cheap”, “affordable” or “expensive” and before this classification, calculate the mean price of each category.
- Related to descriptive statistics this project exposes the count, mean, standard deviation, minimum, maximum, and percentiles of the property's bathrooms

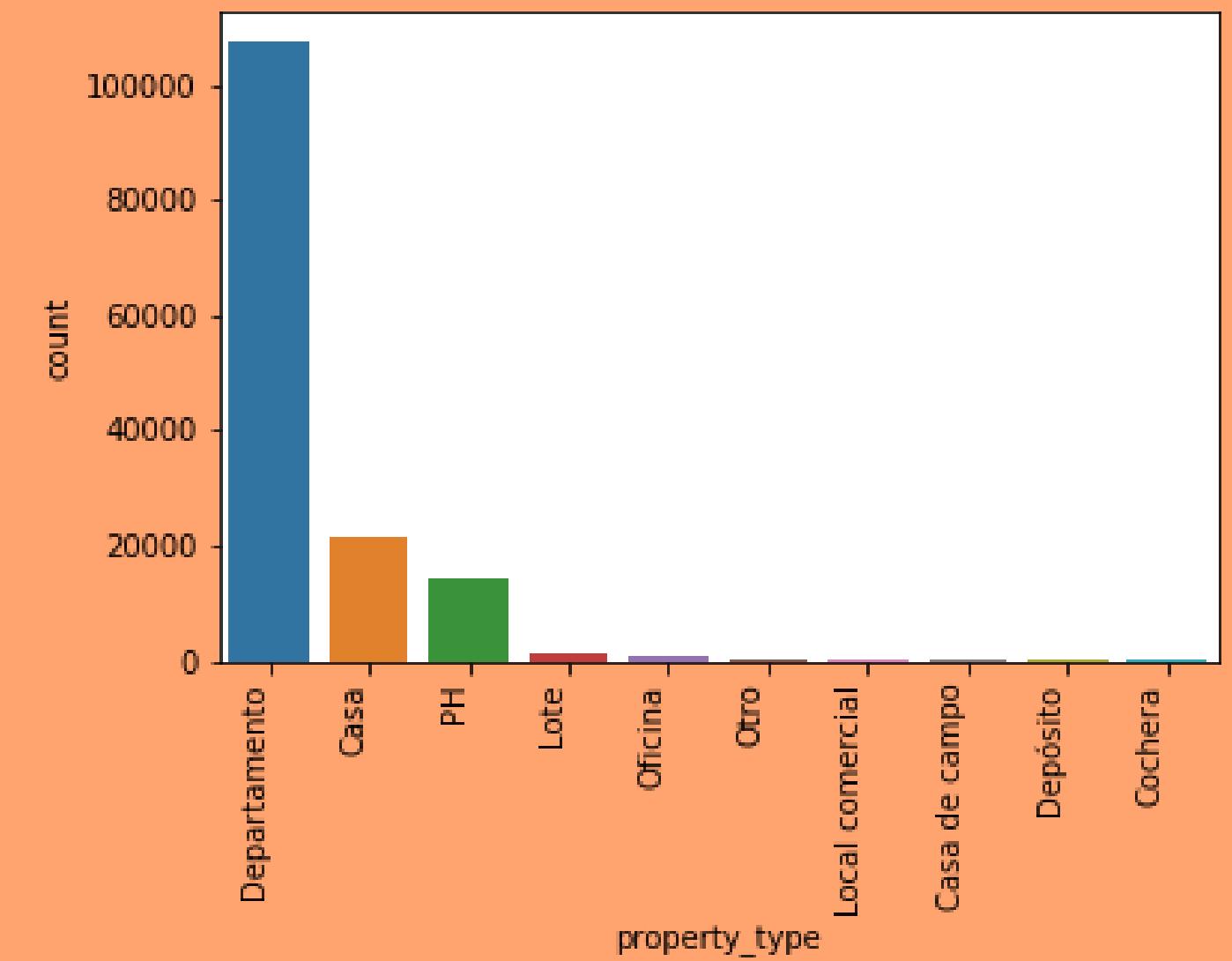
DATA FRAME MANIPULATION

DATA ANALYSIS

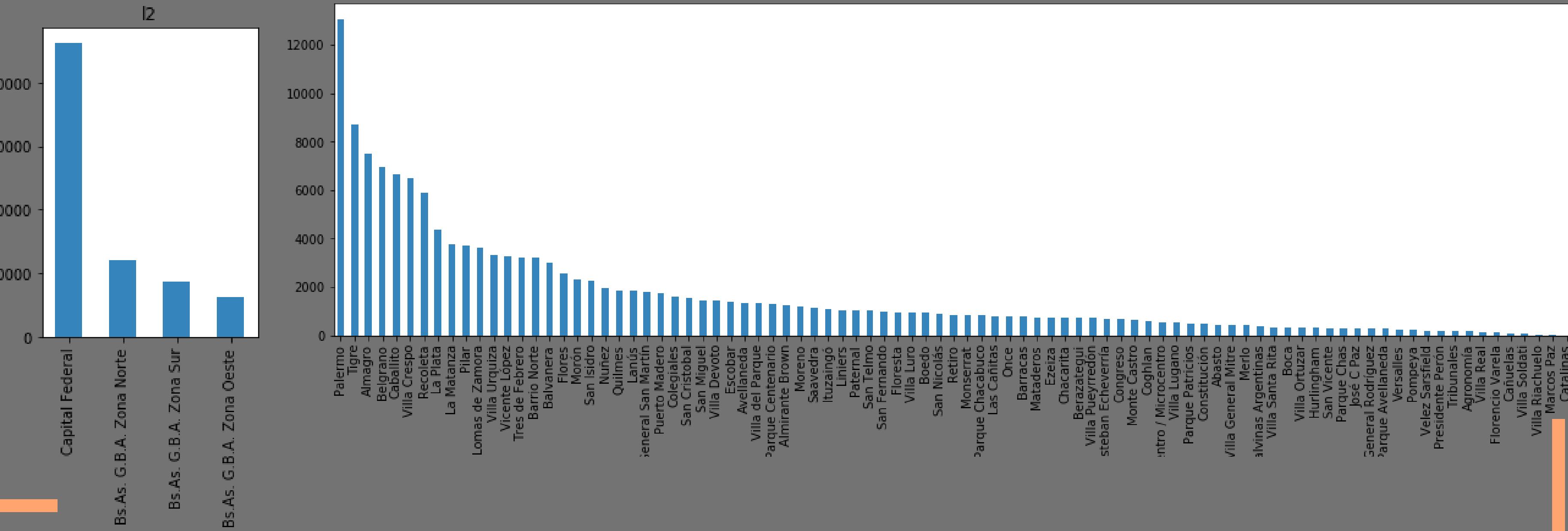
DATA VISUALIZATION

For Data Visualization we imported the matplotlib.pyplot and seaborn. Filtered and analysed the data to present the different graphs like pair plots, general correlation, box plot, scatter plot, line chart, bar graphs, trends, etc. Removing outlier was also one of the task.

PROPERTY TYPE DISTRIBUTION

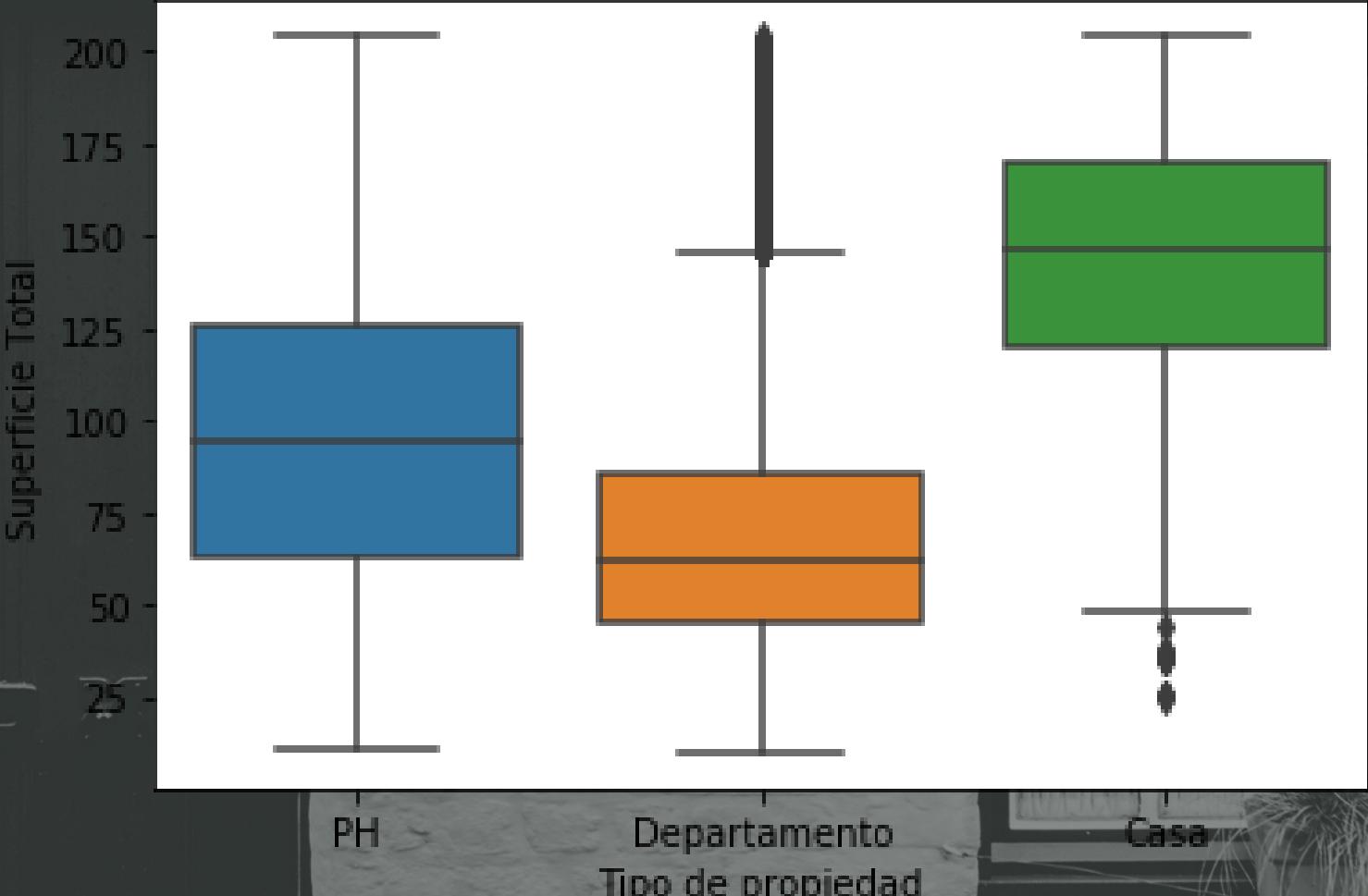


DISTRIBUTION BY REGION



DISTRIBUTION BY PROPERTY TYPE AND SURFACE_TOTAL

Diagrama de caja de la superficie total según el tipo de propiedad



REGRESSION MODEL

For Regression model we are using linear regression, decision tree, nearest neighbors, etc. With the root mean square error we plotted the validation curve. We will train one of closest neighbors and the other of decision trees - to predict the price of "Apartment" type properties in the Autonomous City of Buenos Aires ("Capital Federal"). Keeping the guidelines which corresponds to the "minimum deliverable" in the end we will get a dataset with 81019 instances and 6 columns.

Clustering for Classification

Clustering as a method of finding subgroups within observations is used widely in applications like market segmentation wherein we try and find some structure in the data. Although an unsupervised machine learning technique, the clusters can be used as features in a supervised machine learning model.

KMeans is a clustering algorithm that divides observations into k clusters. Since we can dictate the number of clusters, it can be easily used in a classification where we divide data into clusters that can be equal to or more than the number of classes.

THANK YOU!

IF YOU HAVE ANY DOUBT, DO NOT HESITATE
TO CONTACT US.