

Desafio 1 - Parte 1

Martín Brambilla

Nicolás It

10-Sep-2018

Problema - Problem Statement

Existe un grupo de inversionistas que formaron una alianza con la universidad de Palermo para brindar un servicio de viviendas para alquilar a los estudiantes que vienen del interior.

Estas, tienen ciertas características de forma que cada estudiante pueda alquilar una vivienda de forma individual o de forma compartida con otros estudiantes

Buscamos crear un modelo que logre predecir el precio de un inmueble dadas sus características

Alcance

El alcance de la presente entrega es de realizar un análisis exploratorio de datos de la información provista de [Properati](#) alineada al problema propuesto.

Buscamos tener un entendimiento básico de los datos y de las relaciones existentes entre las variables analizadas.

Para ello, proponemos organizar y preparar los datos, detectar fallos en el lote provisto, transformando y evaluando datos ausentes (missing Data), identificando casos atípicos (outliers) y comprobando algunos supuestos básicos como normalidad y linealidad en las variables.

Estrategia de análisis

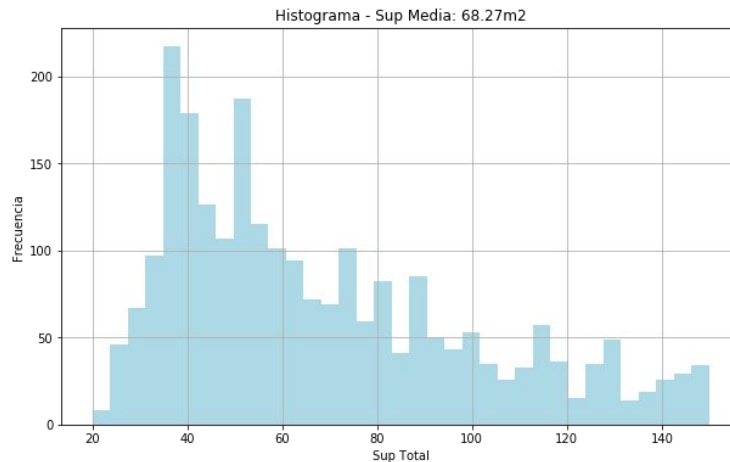
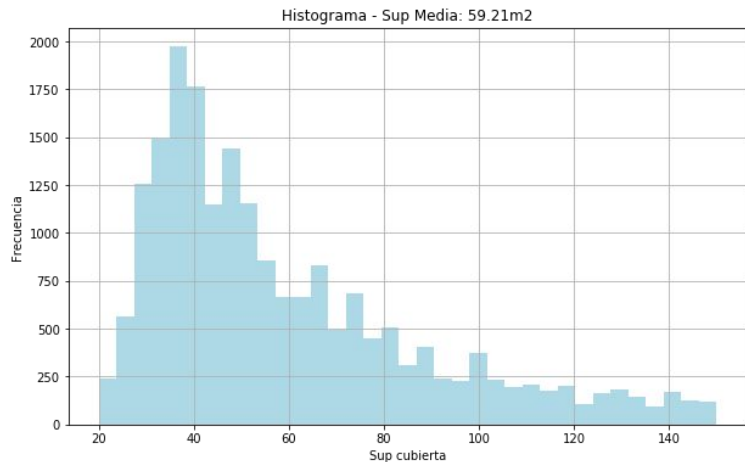
— — —

Identificamos la variables más significativas que se alinean a la propuesta según el siguiente orden:

1. ¿**Cómo** es la propiedad?
2. ¿**Cuál** es su **valor**?
3. ¿**Dónde** se encuentra ubicada?

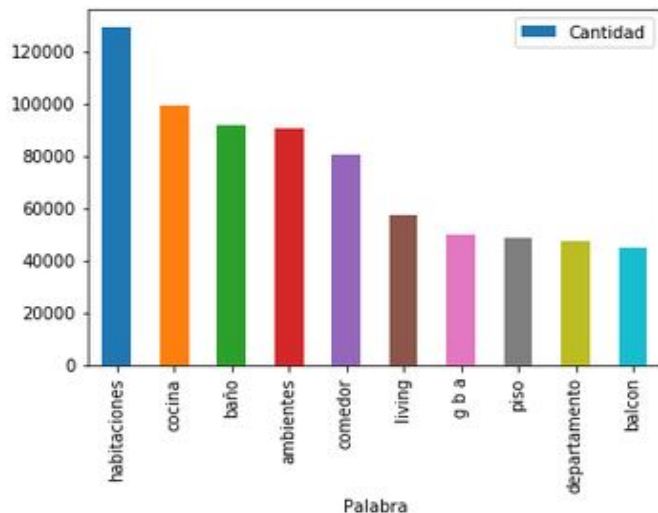
¿Cómo es la propiedad? - Superficie

- Superficie: Consideramos inmuebles de tipo departamento con superficies menores a 150m^2 y mayor a 20m^2
- Buscamos la mejor variable que describa la superficie del inmueble:



¿Cómo es la propiedad? - Atributos

- Se tokenizaron las palabras más usadas.
- Se armó un diccionario de palabras no importantes como artículos y se lo usó como blacklist
- Se normalizaron como flags cada uno de los atributos encontrados dentro de las descripciones



Estas son las 10 palabras mas comunes

```
habitaciones : 129552
cocina : 99244
baño : 91974
ambientes : 90675
comedor : 80391
living : 57268
g b a : 50165
piso : 48641
departamento : 47344
balcon : 45090
```

Relación entre variables

- Precio es directamente proporcional
 - Valor X m²
 - Superficie m²
 - Gastos

