

# SGBD

October 20, 2021

## 1 Trabajo Práctico N°2

### 1.0.1 Procesamiento de tablas

Alumnos:

- Brian Sosa
- Ramiro Ares
- David Veron

### 1.0.2 Montaje de Google Drive para Colab y lectura del data set

```
[ ]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

```
[ ]: !unzip "/content/drive/MyDrive/Colab Notebooks/properati.zip"
```

Archive: /content/drive/MyDrive/Colab Notebooks/properati.zip  
inflating: properati-AR-2018-02-01-properties-sell.csv

```
[ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv("properati-AR-2018-02-01-properties-sell.csv")
df
```

```
[ ]:
image_thumbnail      id ...
0      0de2467efb2476545515f109b506e084728ea6a2 ...
https://thumbs4.properati.com/0/0spuVNRd1jmVHZ...
1      cb23211cab9f6183df65c01f6be24bb48f4dca62 ...
https://thumbs4.properati.com/3/L6iQ8xPFc9vf9b...
2      4f62d7c3f374570841ab10a3cbe6cca938e21484 ...
https://thumbs4.properati.com/8/chd92sZHrIo_n1...
3      1649a3c35359d681abae4a63d9a8545054d2096b ...
https://thumbs4.properati.com/4/VwMxluMwG26qfK...
```

```

4          f7270f3642296d4454458b0a6cfc0fa422d015a7 ...
https://thumbs4.properati.com/7/tLacKGf0tTKun1...
...
...
179658  3cfb3a489574843c63fc4530cc3ff13936d1c334 ...
https://thumbs4.properati.com/0/Jfxnv5jLfqcmWf...
179659  4c3003fb94e85d7b12122e1c976a03009b29e683 ...
https://thumbs4.properati.com/4/oY1DdZNAAdKFtiY...
179660  daaa288b5e16d275f5b10866c1d869862e919fbd ...
https://thumbs4.properati.com/2/m2AsqMYSbiwPba...
179661  d93424b51954e7a5c4704d86d0894474be4efb8d ...
https://thumbs4.properati.com/7/vAXV5zxpSdpSD0...
179662  4ee2810feb951adf0ce176adccd43a86b99642b7 ...
https://thumbs4.properati.com/6/hdtBf5C0ZADjer...

```

```
[179663 rows x 27 columns]
```

## 2 Ejercicio 3.1.1

### 2.1 Calcular el valor medio de los depts 2 ambientes en Capital Federal

```
[ ]: df_dosAmbientes = df.loc[(df['rooms']==2) & (df['state_name'] == 'Capital_Federal')]
df_columnPrice = df_dosAmbientes.loc[:, 'price']
df_columnPrice.mean()
```

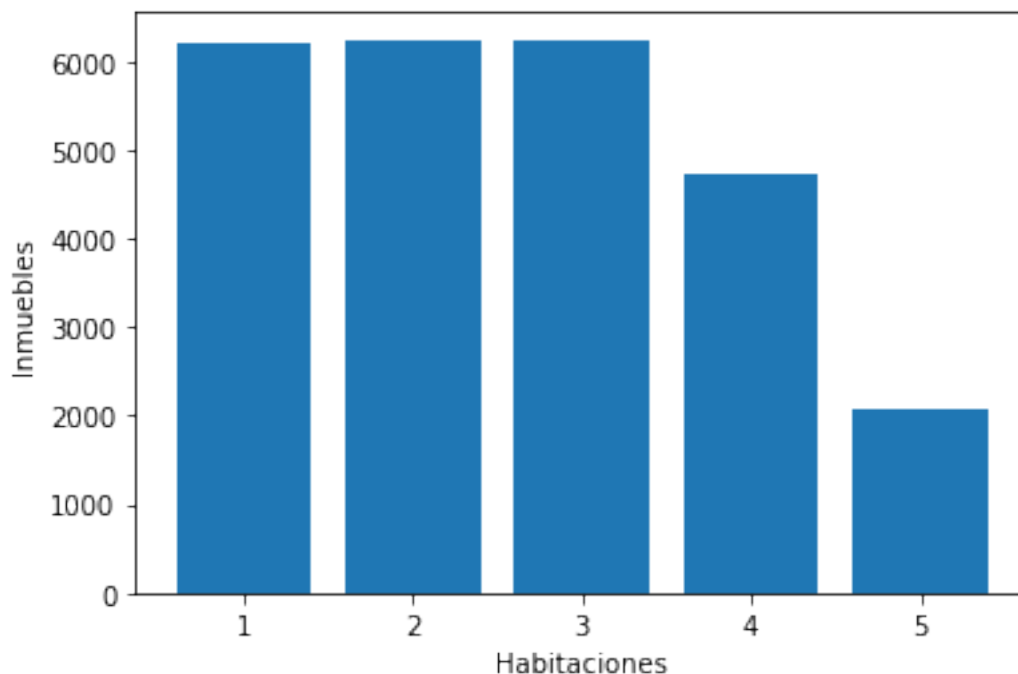
```
[ ]: 294436.56733378564
```

## 3 Ejercicio 3.1.2

### 3.1 Hacer un grafico de barras por cantidad de ambientes en Capital Federal quitando los outliers

```
[ ]: df_caba = df.loc[df['state_name'] == 'Capital Federal']
filtroRooms = df_caba[df['rooms'].between(1, 5)]
df_columnRooms = filtroRooms.loc[:, 'rooms']
x_values = df_columnRooms.unique()
y_values = pd.value_counts(df_columnRooms.values, sort=False)
plt.bar(x_values, y_values)
plt.xlabel('Habitaciones')
plt.ylabel('Inmuebles')
plt.show()
plt.close('all')
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: UserWarning:
Boolean Series key will be reindexed to match DataFrame index.
```

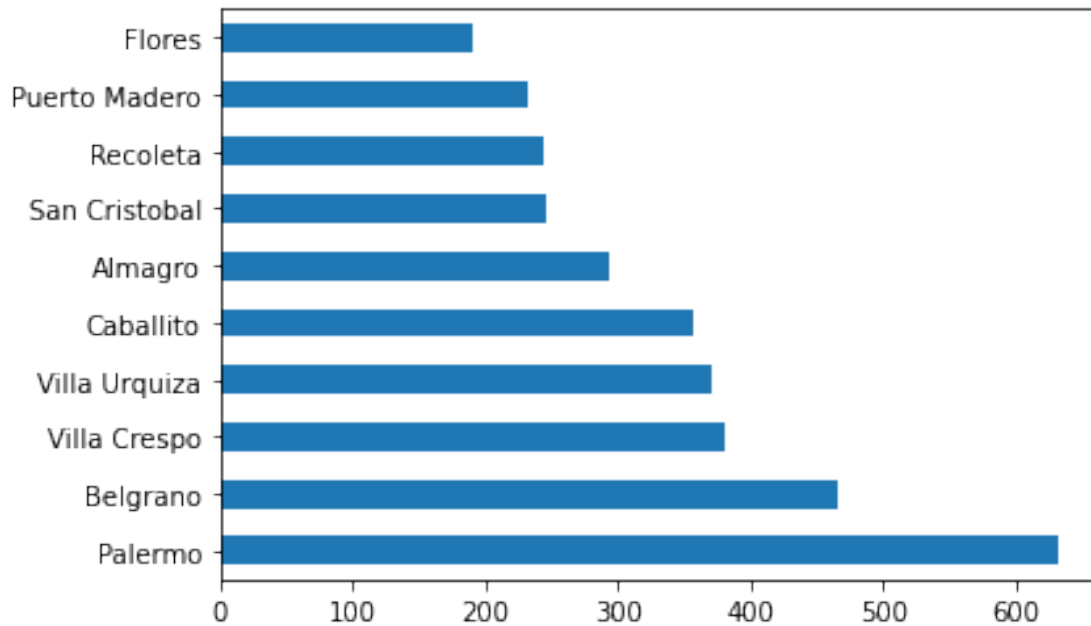


## 4 Ejercicio 3.1.3

4.1 Hacer un grafico de barras horizontal de los 10 barrios con mayor cantidad de publicaciones de deptos. de 2 ambientes en Capital Federal

```
[ ]: df_ambientes = df.loc[(df['rooms']==2) & (df['state_name'] == 'Capital_Federal')]
df_columBarrios = df_ambientes.loc[:, 'place_name']
serie_barrio = pd.Series(df_columBarrios).value_counts().head(10)
serie_barrio.plot.barh()
```

```
[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7fd52d8a1510>
```

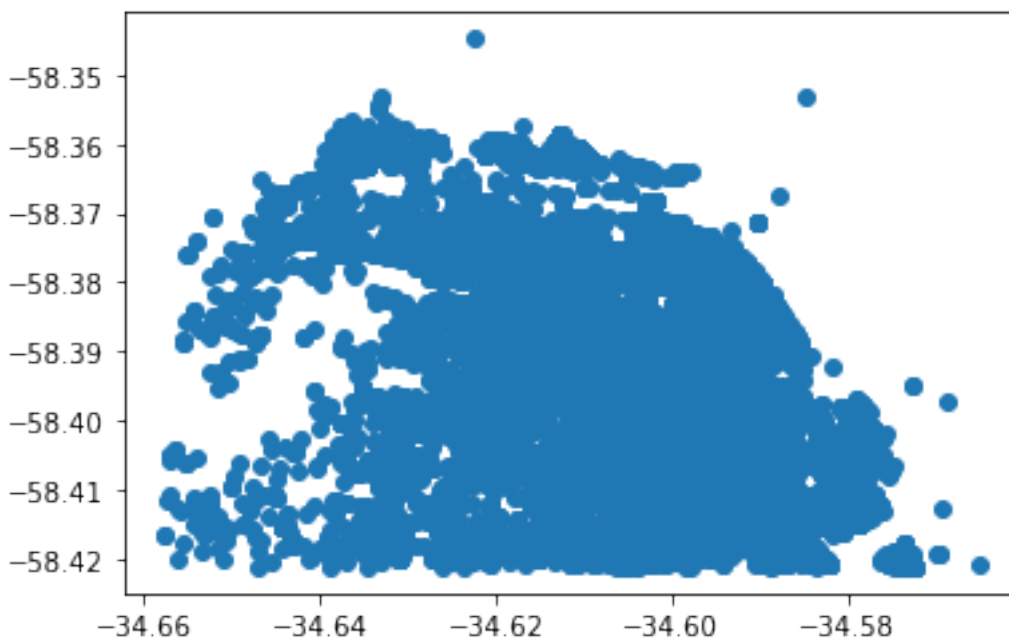


#Ejercicio 3.2 ##### Para aquellas propiedades de Capital Federal que tengan informaci3n geogr3fica se pide escribir un programa para hacer un scatterplot de las propiedades que difieran a lo sumo en 0.05 grados en latitud y longitud respecto al centro geogr3fico de la ciudad. Obs.: obtener las coordenadas del centro de la ciudad de modo aproximado con googlemaps).

```
[ ]: dfCapitalCoordenadas = df[(df['state_name'] == 'Capital Federal') &
    ↳ (df['lat-lon'].notna()) & (df['lat'].notna()) & (df['lon'].notna())]
# Coordenadas Ciudad de Buenos Aires:      Latitud: -34.6083, Longitud: -58.
    ↳ 3712
# Latitud M3nima: -34.6583 ; Latitud M3xima: -34.5583
# Longitud M3nima: -58.4212 ; Longitud M3xima: -58.3212

dfCapitalCoordenadas = dfCapitalCoordenadas.astype({"lat": float, "lon": float})
dfCapitalCoordenadas = dfCapitalCoordenadas[dfCapitalCoordenadas['lat'].
    ↳ between(-34.6583, -34.5583)]
dfCapitalCoordenadas = dfCapitalCoordenadas[dfCapitalCoordenadas['lon'].
    ↳ between(-58.4212, -58.3212)]
plt.scatter(dfCapitalFiltrado['lat'], dfCapitalFiltrado['lon'])
```

```
[ ]: <matplotlib.collections.PathCollection at 0x7f3b2a5d9310>
```



## 5 Ejercicio 3.3

5.1 En el cap. 5 (pag. 96-102) del libro 1 se explican las medidas de tendencia central y de dispersion para describir estadísticamente un conjunto de datos. En este ejercicio la idea es utilizar algunas de esas medidas para realizar especulaciones fundamentadas sobre el conjunto de datos.

Las cinco ciudades con mayor población del país son: 1. Buenos Aires 2. Cordoba 3. Rosario 4. La Plata 5. Mar del Plata

### 6 3.3.1

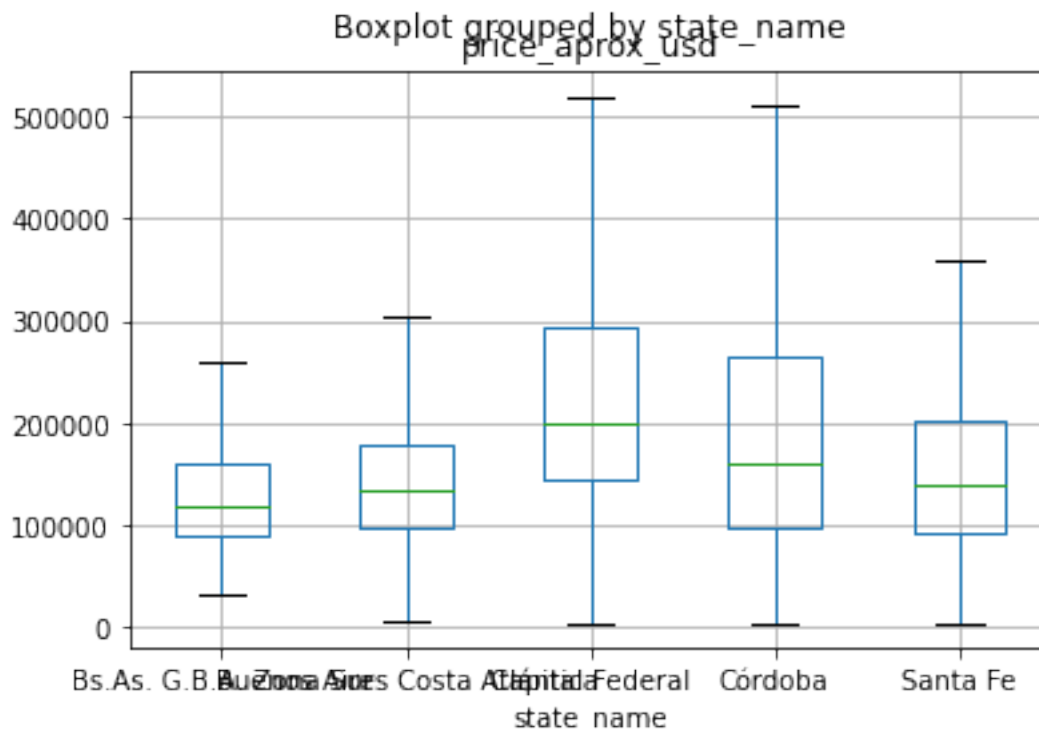
Considerando solamente los deptos. de 3 ambientes, escribir un programa que grafique un boxplot de los precios de esos deptos. de las 5 ciudades mencionadas.

```
[137]: df_tresAmbientes = df.loc[
    (df['rooms']==3) &
    (df['price_aprox_usd'].notna()) &
    ((df['state_name'] == 'Capital Federal') |
    (df['state_name'] == 'Córdoba') |
    (df['place_name'] == 'Rosario') |
    (df['place_name'] == 'La Plata') |
    (df['place_name'] == 'Mar del Plata'))
]
```

```
df_tresAmbientes.boxplot(column=['price_aprox_usd'], by='state_name',
→showfliers=False)
```

```
/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py:83:
VisibleDeprecationWarning: Creating an ndarray from ragged nested sequences
(which is a list-or-tuple of lists-or-tuples-or ndarrays with different lengths
or shapes) is deprecated. If you meant to do this, you must specify
'dtype=object' when creating the ndarray
return array(a, dtype, copy=False, order=order)
```

[137]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f3b0ffce9d0>



## 7 Ejercicio 3.3.2

###Basandose en el gráfico anterior, responder a las siguientes preguntas: 1. Cual es la ciudad con mayor costo de vida? Justificar

- La ciudad con mayor costo de vida es la Ciudad de Buenos Aires ya que tiene una media más alta y dispersión entre sus datos. Los departamentos tienen aproximadamente una mediana de 200000 USD aproximadamente y es la más alta comparada a las de las otras ciudades.

2. Cual es la ciudad más equitativa? Justificar

- La ciudad más equitativa es la ciudad de La Plata, basicamente es la que tiene menor dispersión entre los precios de venta de sus departamentos, además de tener la mediana más baja

si se compara con las demás ciudades.

3. Proponer algunos argumentos por los cuales podría ser incorrecto deducir las dos respuestas anteriores del conjunto de datos que estamos utilizando
- Si nos referimos a costo de vida, esta mal la argumentación ya que no solo hay que basarse en el costo de un departamento, sino que incluye más gastos como por ejemplo las expensas, servicios, alimentos, etc. Lo mismo incluye para la ciudad más equitativa.
  - El data set puede llegar a tener más datos para contemplar más realmente estos datos estadísticos, ya que seguramente hay muchos más departamentos para obtener datos además de los que están en este data set.