

# **Organizacion de datos (75.06)**

## **Análisis exploratorio de datos**

### **Integrantes del grupo:**

- Marinaro, Santiago – Padrón: 97969
- Blanco, Sebastián Ezequiel – Padrón: 98539
- Montiel Maximiliano – Padrón: 93157

**Curso:** 01

**Profesora:** Argerich, Luis Ricardo

**Año:** 2017

**Cuatrimestre:** 2°





# FACULTAD DE INGENIERIA

Universidad de Buenos Aires

## Índice

1. Introducción-----	Pág
2. Instrucciones de instalación-----	Pág
3. Distribución del directorio de carpetas-----	Pág
4. Problemas y errores que acarren los archivos de la empresa properati-----	Pág
5. Hipótesis esperadas para el análisis de los datos-----	Pág
6. Conclusiones-----	Pág

## **1.Introducción**

El objetivo de este trabajo practico es realizar un análisis exploratorio de datos basados en las publicaciones de venta de propiedades, publicadas en los últimos 4 años proporcionada por la empresa Properati.

Los requisitos:

- El análisis se ha hecho en Python Pandas
- El link del Github es:  
[https://github.com/Monty1991/TPDATOS\\_7506\\_2017\\_2\\_ARGERICH/tree/master](https://github.com/Monty1991/TPDATOS_7506_2017_2_ARGERICH/tree/master)

## **2.Instrucciones de instalación**

Para realizar este trabajo es necesario tener instalado los siguientes programas:

- Geopandas
- numpy
- shapely
- fiona
- six
- pyproj
- jupyter notebook

Para poder instalarlo desde la terminal se puede utilizar el siguiente comando:

- `pip install --user <nombre del programa>`

Para ello es necesario tener instalado el programa pip en su ultima versión

## **3.Distribución del directorio de carpetas**

Para el correcto funcionamiento del trabajo, para que los kernel del jupyter sean ejecutados exitosamente se deben tener ciertas carpetas con ciertos archivos en la carpeta donde se encuentra el trabajo práctico.

Para que se ejecute la parte del código que filtra los archivos csv de cada mes de cada años para concatenarlos en uno solo, debe existir una carpeta llamada “properties” donde todos estos archivos se encuentren dentro de ella.

Debe existir la carpeta extras donde se encuentran datasets adicionales donde se evalúan ciertas relaciones de los precios de las propiedades con los subtes, universidades, etc.

En definitiva, al descargar el archivo del repositorio de git debemos tener dentro de esta el archivo TP.ipynb, la carpeta “properties” con todos los datasets de cada mes, la carpeta extras con los datasets adicionales, debemos tener el archivo “barrios.csv” y el archivo “propertiesConCat.csv” que se generara solo, luego de ejecutar el codigo de filtrado de los archivos que los concatena y crea este último.

## **5.Hipótesis esperadas para el análisis de datos**

- Se espera que al agrupar los precios según la fecha en la cual fueron publicados se verá una tendencia creciente a medida que avanzamos en el transcurso de los meses. Se prevee que esto ocurrirá tanto para una análisis discreto para Gran Buenos Aires como para Capital Federal
- Se espera que el precio dependa de qué tipo de propiedades es (casa, PH, apartamento, tienda).
- Se espera que la relación del promedio de los precios en dólares de cada propiedad en función de la cercanía a cierta estación de subte influirá en la suba del mismo, es decir, que al estar mas cerca de la parada de un subte, mas caro será la propiedad.
- Se espera que la relación del promedio de los precios en dólares de cada propiedad en función de la cercanía a los hospitales influirá en la suba del mismo, es decir, que al estar mas cerca de un hospital, mas caro será la propiedad.
- Se espera que la relación del promedio de los precios en dólares de cada propiedad en función de la cercanía a las universidades influirá en la suba del mismo, es decir, que al estar mas cerca de una universidad, mas caro será la propiedad.
- Se espera que la relación del promedio de los precios en dólares de cada propiedad en función de la cercanía a la parada de los colectivos influirá en la suba del mismo, es decir, que al estar mas cerca de la parada de un colectivo, mas caro será la propiedad.
- Se espera que el precio promedio de las propiedades en función de la cantidad de ambientes de las casas será mayor, es decir, que a medida que aumentan las cantidad de ambientes, aumenta el precio promedio.
- Se espera que al agrupar los precios promedio de las propiedades por cada barrio de la ciudad de buenos aires, se verán distribuidos de tal manera que la parte norte de esta localidad será el lugar donde se concentra la zona mas cara de la misma.
- Se espera que al agrupar los precios promedio de las propiedades por cada distrito escolar de la ciudad de buenos aires, se verán distribuidos de tal manera el este de la ciudad sera el lugar con mayor densidad de precios, es decir, donde el precio se muestra mas elevado.

## **Conclusiones**

Las conclusiones están presentes en el archivo de jupyter donde se pueden ver comentarios explicando como se llevo a cabo el análisis y que conclusión sacamos de cada gráfico