

75.06/95.58 Organización de Datos

Segundo Cuatrimestre de 2019

Trabajo Práctico 1: Enunciado

El trabajo práctico 1 de la materia se basa en el análisis de datos del registro histórico de publicaciones de en <https://www.zonaprop.com.ar/>. El set de datos incluye el registro de todas las publicaciones realizadas entre los años 2012 y 2016 en México.

Los datos se pueden obtener haciendo clic [aquí](#).

El dataset consta de propiedades en venta en México entre los años 2012 y 2016, valuadas en pesos mexicanos. El archivo *train.csv* tiene 240K filas y 22 columnas. El equipo de Navent preparó [un tutorial](#), en formato de jupyter notebook.

- *id*: Un id numérico para identificar la propiedad
- *titulo*: El título de la propiedad publicada
- *descripcion*: La descripción de la propiedad publicada
- *direccion*: La dirección de la propiedad
- *ciudad*: La ciudad de la propiedad
- *provincia*: La provincia donde está localizada la propiedad
- *lat*: Latitud
- *lng*: Longitud
- *tipodepropiedad*: El tipo de propiedad (Casa, departamento, etc)
- *metrostotales*: Metros totales de la propiedad
- *metroscubiertos*: Metros cubiertos de la propiedad
- *antiguedad*: Antigüedad de la propiedad
- *habitaciones*: Cantidad de habitaciones
- *garages*: Cantidad de garages
- *banos*: Cantidad de baños
- *fecha*: Fecha de publicación
- *gimnasio*: Si el edificio o la propiedad tiene un gimnasio
- *usosmultiples*: Si el edificio o la propiedad tiene un SUM
- *piscina*: Si el edificio o la propiedad tiene un piscina
- *escuelascercanas*: Si la propiedad tiene escuelas cerca
- *centroscomercialescercanos*: Si la propiedad tiene centros comerciales cerca
- *precio*: Valor de publicación de la propiedad en pesos mexicanos

El objetivo del primer TP es realizar un análisis exploratorio del set de datos. Queremos ver qué cosas podemos descubrir sobre los datos que puedan resultar interesantes. Estas cosas pueden estar relacionadas al objetivo del TP2 (predecir el valor del inmueble) o no, ambas son de interés.

Los requisitos de la primera entrega son los siguientes:

- El análisis debe estar hecho en Python Pandas o R.
- El análisis debe entregarse en formato papel en una carpeta en donde se incluya el reporte completo y todas las visualizaciones generadas. Es altamente recomendable que las visualizaciones se impriman en color. En el informe no va código.
- Informar el link a un repositorio Github en donde pueda bajarse el código completo para generar el análisis.
- Agregar en Kaggle un kernel con el análisis exploratorio realizado.

La evaluación del TP se realizará en base al siguiente criterio:

- Originalidad del análisis exploratorio.
- Calidad del reporte. ¿Está bien escrito? ¿Es claro y preciso?
- Calidad del análisis exploratorio: qué tipo de preguntas se hacen y de qué forma se responden, ¿es la respuesta clara y concisa con respecto a la pregunta formulada?
- Calidad de las visualizaciones presentadas.
 - ¿Tienen todos los ejes su rótulo?
 - ¿Tiene cada visualización un título?
 - ¿Es entendible la visualización sin tener que leer la explicación?
 - ¿El tipo de plot elegido es adecuado para lo que se quiere visualizar?
 - ¿Es una visualización interesante?
 - ¿El uso del color es adecuado?
 - ¿Hay un exceso o falta de elementos visuales en la visualización elegida?
 - ¿La visualización es consistente con los datos?
 - Conclusiones presentadas.
 - ¿Presenta el grupo un listado de "insights" aprendidos sobre los datos en base al análisis realizado? ¿Es interesante?
 - ¿Descubrieron cosas que pueden ser de interés a las personas en busca de propiedades o que se encuentre en el mercado inmobiliario?
 - ¿El análisis realiza un aporte a Navent?

El grupo que realice el mejor análisis exploratorio obtendrá 10 puntos para cada uno de sus integrantes que podrán ser usados en el parcial además de ser publicado en el repositorio de la materia como ejemplo para los siguientes cuatrimestres.