## **Reto de Data Science – Comparables Inmobiliarios en Cuauhtémoc, CDMX**

### **Objetivo del Reto**

El objetivo de este reto es desarrollar una función get\_comparables que recibe a lo mucho 7 inputs, devuelve una lista con los url\_ad de las 5 propiedades más similares de una tabla con propiedades.

Para realizar esto se te compartirá un conjunto de propiedades dentro del municipio de **Cuauhtémoc, Ciudad de México**.

### **Diccionario**

price: Precio de la propieadad

url\_ad: link al anuncio de la propiedad (puede que te salgan expirados por que será una tabla estática)

num\_bedrooms: Número de dormitorios

num\_bathrooms: Numero de baños

construction\_surface: Área de construcción

terrain\_surface: Área de terreno

built\_year: Año de construcción ó remodelación de la propiedad

latitude: latitud

longitude: longitud

neighborhood: colonia

id\_neighborhood: id de la colonia

conservation\_status: estado de conservación (valores del 0-1)

property\_type: casa o departamento

listing\_type: renta o venta (solo te proporciono de venta)

has\_gym: 1 si tiene gimnasio, 0 no tiene

has\_garden: 1 si tiene jardín, 0 no tiene



### **Expectativas**

Tu entregable deberá ser un notebook con comentarios claros para comprensión del proceso y párrafos explicativos (opcional: estas explicaciones se pueden incluir en un formato de texto independiente, con que sea clara la relación entre ambos.. Con solo leer este (estos) archivo, debería ser claro.

* + ¿Cómo definiste “comparabilidad” en tus palabras?
  + ¿Qué análisis/visualización te ayudó a llegar a la definición?
  + ¿Qué transformaciones/feature engineering relevante tuviste que aplicar y por qué?
  + Limitaciones o mejoras posibles (roadmap de mejora).
  + Explicar qué podría escalar mal (por ejemplo, si crece el número de propiedades).

El notebook debe al menos incluir:

* Análisis Exploratorio de los datos
* Fase de experimentación (Nos interesa ver cómo abordaste el problema y que intentaste aunque no haya resultado)
* La construcción y definición de tu algoritmo final.

### **Tiempo estimado:**

1. Tendrás 5 días para completarlo a partir del inicio del reto .
   * Si tienes preguntas concretas, mandalas a [Juan Manuel Hernández](mailto:juanmanuel.hernandez@dd3tech.mx) o a [Christopher Román Jaimes](mailto:christopher.roman@dd3tech.mx)

### **Aclaraciones:**

* + Estas 7 variables de input deberán ser datos categóricos o escalares (número de cuartos, tamaño de construcción, raíz cuadrada del estado de conservación entre 28, lo que sea) no es válido contar un arreglo de datos como uno de estos inputs.
  + Tú deberás definir qué significa “comparable”: puedes usar algoritmos predefinidos o alguna métrica compuesta.
  + Librerías que te pueden funcionar: pandas, geopandas, folium, matplotlib, numpy, shapely, scikit-learn.
  + Recuerda generar un algoritmo eficiente y escalable, aunque esto no es un puesto de software engineering, las reglas y métodos deben de funcionar rápido y ante muchos datos.