

## Proyecto No.2

En este proyecto competiran en Grupos (los mismos grupos del proyecto no. 1) con sus demas compañeros, el objetivo es encontrar una funcion generada a partir de una regresion lineal entrenada, que nos permita estimar el precio de renta de ciertas propiedades.

Deben de inscribirse en el siguiente link de kaggle. Solo debe de inscribirse un integrante por grupo (los mismos grupos del proyecto 1).

<https://www.kaggle.com/t/46597f792b5441fba8dd21702807da68>

En la competencia solo debe de existir un integrante que suba su trabajo y deben de identificarse de la siguiente manera: Grupo 1 (G1) .

Se entregaran dos datasets:

**Dataset: house\_rent\_result90.csv (DATASET DE ENTRENAMIENTO)**

**Rent:** Renta en unidades monetarias de casas o apartamentos. (variable objetivo)

Caracteristicas:

BHK: Number of Bedrooms, Hall, Kitchen.

.Size: Size of the Houses/Apartments/Flats in Square Feet.

Floor: Houses/Apartments/Flats situated in which Floor and Total Number of Floors (Example: Ground out of 2, 3 out of 5, etc.)

Area Type: Size of the Houses/Apartments/Flats calculated on either Super Area or Carpet Area or Build Area.

Area Locality: Locality of the Houses/Apartments/Flats.

City: City where the Houses/Apartments/Flats are Located.

Furnishing Status: Furnishing Status of the Houses/Apartments/Flats, either it is Furnished or Semi-Furnished or Unfurnished.

Tenant Preferred: Type of Tenant Preferred by the Owner or Agent.

Bathroom: Number of Bathrooms.

Universidad Rafael Landívar  
Facultad de Ingeniería  
Análisis de Datos  
Msc. Ing. Jairo Salazar

Point of Contact: Whom should you contact for more information regarding the Houses/Apartments/Flat

Con este dataset se deben de correr una serie de experimentos que nos permita estimar una regresion lineal ideal para poder efectuar nuestra prueba con el set de pruebas de la competencia.

### **Dataset: house\_rent\_result10.csv (DATASET DE TEST PARA LA COMPETENCIA)**

Características:

BHK: Number of Bedrooms, Hall, Kitchen.

.Size: Size of the Houses/Apartments/Flats in Square Feet.

Floor: Houses/Apartments/Flats situated in which Floor and Total Number of Floors (Example: Ground out of 2, 3 out of 5, etc.)

Area Type: Size of the Houses/Apartments/Flats calculated on either Super Area or Carpet Area or Build Area.

Area Locality: Locality of the Houses/Apartments/Flats.

City: City where the Houses/Apartments/Flats are Located.

Furnishing Status: Furnishing Status of the Houses/Apartments/Flats, either it is Furnished or Semi-Furnished or Unfurnished.

Tenant Preferred: Type of Tenant Preferred by the Owner or Agent.

Bathroom: Number of Bathrooms.

Point of Contact: Whom should you contact for more information regarding the Houses/Apartments/Flat

Rownum : numero de registro.

Al terminar de efectuar el entrenamiento y estimar la Funcion del modelo (con el intercepto y las pendientes de cada feature seleccionado), estimar la Renta de acuerdo a su modelo con el dataset de TEST. Y generar un csv con dos columnas: Rownum,Rent. La columna Rent es la estimada de acuerdo a su modelo.

**Recomendaciones:**

1. Efectue un análisis de la información, determinando que características se deben transformar o modificar para poder entrenar el modelo.
2. Efectue el proceso de feature engineering de acuerdo al análisis del punto anterior.
3. Diseñe su experimento seleccionando o transformando características de la propiedad que nos ayuden a poder tener un modelo que explique los datos con un nivel aceptable.
4. Entrene el modelo de acuerdo a las características de la propiedad que considera relevantes y transformadas de acuerdo a su análisis.
5. La función `lm`, nos devuelve el intercepto y la pendiente de cada característica. Esta información debe guardarse pues nos servirá para efectuar una estimación final.
6. Guarde los resultados de cada experimento en conjunto con los resultados obtenidos de experimentos anteriores, con el fin de evitar problemas de repetir experimentos no correctos. El orden es importante.

**Entregables:**

1. Código en un R notebook de las transformaciones de datos, entrenamiento del modelo y generación del archivo para subir al kaggle.
2. PDF justificando las decisiones del modelo (detallar todas sus decisiones).
3. Listado de experimentos efectuados.
4. La entrega es en grupos.

**Calificación:**

1. **Posición final en la competencia (25 puntos)**
2. **Entrega final (75 puntos)**
  - a. **Código en R**
  - b. **PDF de la presentación de su trabajo**