Problem Set 3 – Big Data

https://github.com/BigData-Gomez-Ortiz-Vanegas

Introducción

Este ejercicio propone utilizar el algoritmo *superlearners.* La ventaja principal de este modelo es que permite ponderar varios algoritmos individuales para crear uno nuevo que se desempeñe al menos tan bien como cualquiera de ellos. Así, *superlearners* le asigna un peso a cada algoritmo estadístico individual. Dado el poco seguimiento que tenemos los humanos sobre este comando, una desventaja a reconocer es la posibilidad de cometer errores que pasen desapercibidos. Los algoritmos individuales que incluimos son: *Random Forests*, *GXBoost* y un modelo lineal con el propósito de suavizar la predicción. Las variables que escogimos para ejecutar el algoritmo cuentan con el soporte de la intuición económica y se explican a profundidad a continuación.

Datos

Imputación: se utilizó la variable de *área cubierta* para completar la variable de *área total* en los casos en los que el dato no existiera para la última. Después, se capturó información disponible sobre el *área total* y el *número de baños* en la descripción del anuncio para completar los valores faltantes en dichas variables. También, se utilizó la capa de manzanas para agregar la mediana de *área total* de la manzana donde persistieran valores faltantes.

Selección de la muestra: como grupo, concluimos que es interesante observar los datos a nivel de la localidad (Chapinero y El Poblado) porque los apartamentos deben ser más similares. Sin embargo, al revisar los datos, constatamos que esto no era posible para El Poblado en tanto que casi todos los datos de entrenamiento se encontraban fuera de dicha localidad. En vista de esto, se utilizará la muestra a nivel de Chapinero y a nivel de Medellín; consideramos que es un ejercicio interesante para contrastar.

Nuevas variables: hubo dos estrategias principales para crear las nuevas variables - recurrir a la descripción de los anuncios y utilizar fuentes externas de información (Open Street Map). Se capturó la información disponible en los anuncios para identificar si se menciona que el apartamento *cuenta con terraza o no* y si *cuenta con garaje o no*. Estas variables son interesantes en tanto que la existencia de terraza o garaje pueden significar un aumento en el valor del apartamento. Para la creación de otras variables se obtuvo en Open Street Map las *estaciones de transporte masivo*, *los cerros* y los *campos de golf* disponibles en el área que estamos analizando. Se consideró que una facilidad de transporte público puede valorizar un apartamento, así que se creó la variable de *distancia mínima a la estación de transporte masivo más cercana.* Consideramos que, para Bogotá, es bastante apetecida la zona cercana a los cerros por varias razones (como una mejor calidad del aire) así que agregamos *la distancia mínima a los cerros orientales* para los datos de Bogotá. Por último, nos pareció interesante ver un campo de golf en El Poblado, Medellín y creemos que puede implicar un mayor precio estar cerca de él por la vista y por un "estatus" social. Agregamos la *distancia mínima al campo de golf* para los datos de Medellín.

Tabla 1. Estadísticas descriptivas de variables seleccionadas

Las tablas más completas con las estadísticas descriptivas se pueden observar en el Anexo 1

	Chapinero Train	Chapinero Test	Medellín Train	Medellín Test
Precio	1409463315.64712		415552398.354574	
Habitaciones	2.62282943368943	1.91172761664565	3.15904467690441	3.02375205175244
Área total	138.960904412101	80.0573211963589	98.6922806356912	139.499098495876
Terraza	0.545052625455512	0.383354350567465	0.58993917157478	0.501496572366515
Garaje	0.470313259620299	0.436317780580076	0.46134299643495	0.58993917157478
Distancia a transporte	976.030085594007	309.259247483354	888.882010696326	2678.1595573365

Distancia a cerros/Distancia a golf	1604.8417816846	1867.96357768742	4039.31184319051	1573.22478128512

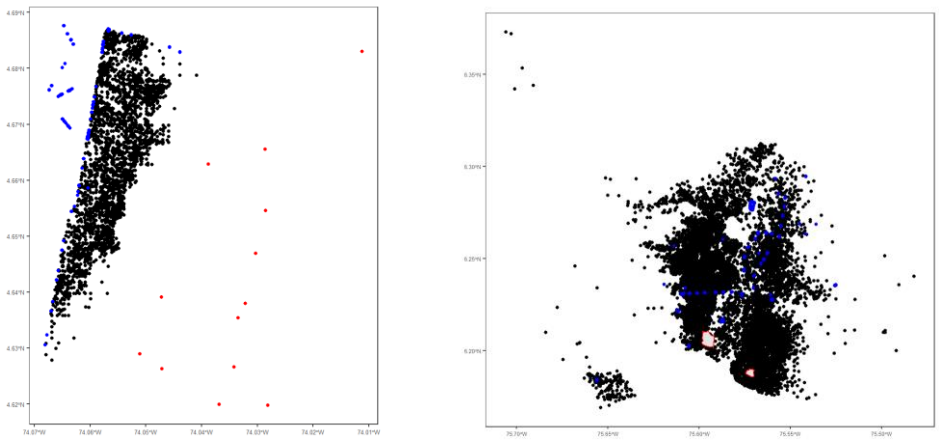
En cuanto a las variables seleccionadas, se ha establecido que diversos factores pueden afectar positiva o negativamente el precio de un apartamento. En términos de localización, se ha establecido que una mayor distancia al centro y de zonas de accesibilidad como paradas de buses afecta negativamente el precio (Chica-Olmo et al., 2020; Deboosere et al., 2019; Dudas et al., 2020). Por otro lado, se ha encontrado una relación positiva entre el número de habitaciones y el precio, así como en el metraje o la superficie de la vivienda (Chica-Olmo et al., 2020; Deboosere et al., 2019; Dudas et al., 2020). Otras investigaciones refieren que la presencia de un mayor número de baños aumenta el valor a pagar y los inmuebles que cuentan con terraza tienen un precio de venta superior al 25% dependiendo de la ubicación en comparación con los que no tienen (Chattopadhyay & Mitra, 2019; Chica-Olmo et al., 2020).

El desarrollo sostenible y los niveles de conciencia sobre el medio ambiente, niveles de polución en el aire, entre otros han llevado a los Bogotanos a buscar un sin número de servicios ecoamigables que brinden un valor agregado. Por ejemplo, la vivienda cerca a los cerros orientales, especialmente en la localidad de Chapinero. Utilizando la metodología de precios hedónicos desarrollada por Rosen (1974), Garzón y Cardozo (2019) afirman que el precio de las viviendas "es directamente proporcional a la cercanía de los cerros", generando un precio comparativamente más elevado en el mercado.

En el anexo 2 se puede observar el delineado de Chapinero, Bogotá y Medellín. A continuación, se muestran las variables espaciales relevantes para cada sector. Para el caso de Chapinero, Bogotá, se evidencia los apartamentos en venta (círculos negros), las estaciones de transporte masivo (círculos azules) y los cerros orientales (círculos rojos). Para el caso de Medellín, se observa los apartamentos en venta (círculos negros), las estaciones de transporte masivo (círculos azules) y el campo de golf (polígono rojo).

Figura 1. Mapas de Chapinero, Bogotá y El Poblado, Medellín con datos espaciales disponibles

Chapinero, Bogotá **Medellín**



Modelo y resultados

Variables: Las variables tenidas en cuenta para el modelo son: el área total (*surface2*), si tiene terraza o no (*tiene_terraza*), si tiene garaje o no (*tiene_garaje*), el número de habitaciones (*bedrooms*), la distancia mínima a las estaciones de transporte masivo (*dist_*bus), la distancia mínima a los cerros orientales (en el caso de Bogotá – *dist_east*) y la distancia mínima al campo de golf (en el caso de El Poblado - *dist_golf)*. Se incluyeron estas variables ya que la intuición económica, y en parte la literatura

expuesta en la sección anterior, nos indicaron que estos factores podrían implicar una variación importante en el precio de los apartamentos. Además, se tuvo presente una función lineal y cuadrática (disponibles en Anexos 3 y 4) de dichas variables, finalmente se conservó la versión cuadrática de la función. Para el modelo final se eligió la función cuadrática por su menor error cuadrático medio en los resultados.

Modelo y evaluación: Tanto para Chapinero como para Medellín se ejecutaron los algoritmos de probabilidad lineal, *Random Forest*, *XGBoost* y *SuperLearner*. El modelo elegido fue *SuperLearner* ya que funciona por lo menos tan bien como cualquiera de los algoritmos y mostró buen desempeño con respecto a los demás (resultados disponibles en Anexos 3-7). Así, para entrenar el modelo final se utilizó la función *SuperLearner* disponible en R y se introdujo como algoritmos *RandomForest, GXBoost* y un modelo de probabilidad lineal para suavizar la predicción. A continuación, se puede ver el error mínimo y los coeficientes de ponderación que arrojó el modelo.

Tabla 2. Error mínimo y coeficientes de ponderación para modelo Chapinero y Medellín

	Chapinero			**Medellín**	
	Risk	Coef		Risk	Coef
SL.lm_All	5.358729e+17	0.01166931	SL.lm_All	1.213429e+17	0.04383829
SL.rpart_All	5.214751e+17	0.00000000	SL.rpart_All	1.032308e+17	0.05110621
SL.xgboost_All	3.523261e+17	0.98833069	SL.xgboost_All	7.384755e+16	0.90505550

Conclusiones

Después de considerar los modelos de probabilidad lineal, *Random Trees*, *XGBoost* y *SuperLearner*, este último fue el elegido por su mejor desempeño. Esto tiene sentido ya que, como se mencionó, *SuperLearner* le da un peso de relevancia a cada uno de los algoritmos introducidos y se desempeña por lo menos tan bien como cualquiera de ellos. Con base en la tabla 2, se evidencia que el modelo para chapinero le da un peso de 99% a *XGBoost* y 1% al modelo de probabilidad lineal, mientras que el modelo para Medellín otorga un peso de 91% a *GXBoost*, de 5% a *RandomForests* y de 4% al modelo de probabilidad lineal.

Anexos

Anexo 1. Estadísticas descriptivas

Chapinero Train Stats							
Stat	**price**	**bedrooms**	**surface2**	**tiene_terraza**	**tiene_garaje**	**dist_chapi_bus**	**dist_east**
# Values	240663	240663	240543	240663	240663	240663	240663
# Null	0	541	0	109489	127476	0	0
# NA	0	0	120	0	0	0	0
Median	1.2e+09	3	106	1	0	1118.92364745359	1416.00894978402
Mean	1409463315.64712	2.62282943368943	138.960904412101	0.545052625455512	0.470313259620299	976.030085594007	1604.8417816846
Std. Dev	853466930.238821	0.920068950620243	55.3171553151136	0.497967158863804	0.499118956347023	443.86294523022	578.846919439646

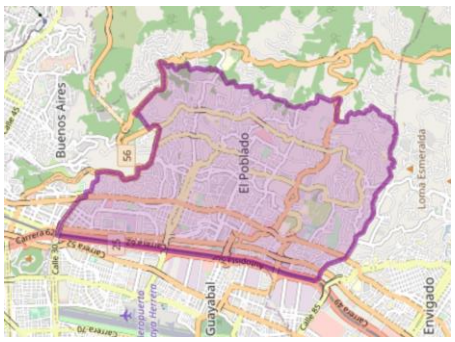
Chapinero Test Stats						
Stat	**bedrooms**	**surface2**	**tiene_terraza**	**tiene_garaje**	**dist_chapi_bus**	**dist_east**
# Values	793	769	793	793	793	793
# Null	3	0	489	447	0	0
# NA	0	24	0	0	0	0
Median	1	78	0	0	305.81985370436	1840.589046744428
Mean	1.911727616645 65	80.05732119635 89	0.383354350567465	0.436317780580076	309.259247483354	1867.963577687 42
Std. Dev	1.272187397437 37	33.88562417349 23	0.486510297402999	0.496240980783741	140.74915788086	232.232053622339

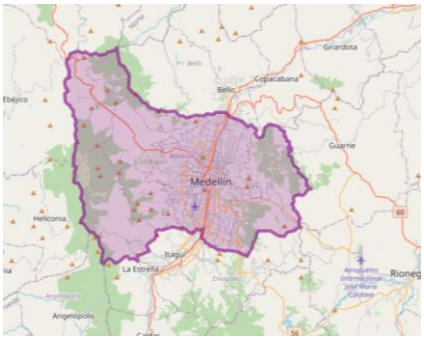
Medellín Train Stats							
Stat	**price**	**bedrooms**	**surface2**	**tiene_terraza**	**tiene_garaje**	**dist_med_bus**	**dist_golf**
# Values	290038	290038	289197	290038	290038	290038	290038
# Null	0	42	2	158375	156231	0	117
# NA	0	0	841	0	0	0	0
Median	308796000	3	100	0	0	922.022570293893	4018.75409788119
Mean	415552398.354574	3.15904467690441	98.6922806356912	0.453950861611237	0.461342996 43495	888.882010696326	4039.31184319051
Std. Dev	430838691.414588	1.09909952385458	161.604055044151	0.497875819358028	0.498504255630824	572.946270294205	1988.22157077419

Medellín Test Stats						
Stat	**bedrooms**	**surface2**	**tiene_terraza**	**tiene_garaje**	**dist_med_bus**	**dist_golf**
# Values	10357	10305	10357	10357	10357	10357
# Null	13	0	5163	4247	0	11
# NA	0	52	0	0	0	0
Median	3	120	1	1	2626.45058673266	1426.30684730773

Mean	3.02375205175244	139.499098495876	0.501496572366515	0.589939171574 78	2678.159557336 5	1573.224781285 12
Std. Dev	0.911079234904373	60.6717777463658	0.500021900170085	0.491868178390216	1027.540870682 53	909.0743855040 15

Anexo 1. Perímetro de Chapinero, Bogotá y El Poblado, Medellín





Anexo 2. Variables y formas funcionales tenidas en cuenta

Lineal

- Bogotá

$$Precio_a = \acute{A}reaTotal_a + Habitaciones_a + Terraza_a + Garaje_a + Distancia\ mínima\ a\ transporte\ masivo_a + Distancia\ mínima\ a\ cerros\ orientales_a + u$$

- Medellín

$$Precio_a = \acute{A}reaTotal_a + Habitaciones_a + Terraza_a + Garaje_a + Distancia\ mínima\ a\ transporte\ masivo_a + Distancia\ mínima\ a\ campo\ de\ golf_a + u$$

Cuadrática

- Bogotá

$$Precio_a = \acute{A}rea\ Total_a^2 + Habitaciones_a + Terraza_a + Garaje_a + Diastancia\ mínima\ a\ transporte\ masivo_a^2 + DIstamcia\ mínima\ a\ cerros\ orientales_a^2 + u$$

- Medellín
  - $Precio_a = \acute{A}rea\ Total_a^2 + Habitaciones_a + Terraza_a + Garaje_a + Diastancia\ mínima\ a\ transporte\ masivo_a^2 + DIstamcia\ mínima\ a\ campo\ de\ golf\ orientales_a^2 + u$

Anexo 3. Resultados modelo lineal con forma funcional lineal

- Bogotá

```
Chapinero OLS
================================================
                        Dependent variable:
                    ---------------------------
                               price
------------------------------------------------
dist_east                   -56,726.310**
                             (24,896.430)

dist_chapi_bus              395,499.400***
```

	(32,990.240)
bedrooms	359,632,926.000***
	(6,361,317.000)
tiene_terraza1	266,894,535.000***
	(13,046,137.000)
tiene_garaje1	-51,321,618.000***
	(12,840,069.000)
surface2	2,622,443.000***
	(114,666.900)
Constant	-291,968,062.000***
	(71,444,564.000)
Observations	13,365
R2	0.351
Adjusted R2	0.351
Residual Std. Error	**729,021,896.000 (df = 13358)**
F Statistic	1,203.675*** (df = 6; 13358)
Note:	*p<0.1; **p<0.05; ***p<0.01

- Medellín

Medellín OLS

	Dependent variable:
	price
dist_golf	-32,837.110***
	(1,090.284)
dist_med_bus	110,568.400***
	(3,514.413)
bedrooms	134,379,756.000***
	(2,246,766.000)
tiene_terraza1	106,292,344.000***
	(6,360,533.000)
tiene_garaje1	31,075,228.000***
	(4,859,700.000)
surface2	49,284.310***
	(5,695.814)
Constant	-12,806,009.000
	(9,330,555.000)
Observations	20,778
R2	0.216
Adjusted R2	0.216
Residual Std. Error	**348,083,111.000 (df = 20771)**
F Statistic	953.825*** (df = 6; 20771)
Note:	*p<0.1; **p<0.05; ***p<0.01

Anexo 4. Resultados modelo lineal con forma funcional cuadrática

- Bogotá

```
Chapinero OLS
================================================
                        Dependent variable:
                    ----------------------------
                                price
------------------------------------------------
dist_east2                    -37.283***
                               (4.700)

dist_chapi_bus2               166.203***
                               (12.729)

bedrooms                   359,953,754.000***
                            (6,365,374.000)

tiene_terraza1             270,089,913.000***
                            (13,053,890.000)

tiene_garaje1              -51,445,523.000***
                            (12,850,243.000)

surface2                    2,635,714.000***
                             (114,736.500)

Constant                   -86,843,142.000***
                            (33,489,283.000)

------------------------------------------------
Observations                      13,365
R2                                0.350
Adjusted R2                       0.350
Residual Std. Error 729,577,622.000 (df = 13358)
F Statistic         1,198.452*** (df = 6; 13358)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

- Medellín

```
Medellín OLS
================================================
                        Dependent variable:
                    ----------------------------
                                price
------------------------------------------------
dist_golf2                    -3.627***
                               (0.105)

dist_med_bus2                 16.482***
                               (0.542)

bedrooms                   133,434,549.000***
                            (2,240,665.000)

tiene_terraza1             103,131,321.000***
                            (6,343,925.000)

tiene_garaje1              30,547,622.000***
                            (4,844,049.000)

surface2                    51,144.050***
                             (5,678.414)

Constant                   21,589,800.000***
                            (8,239,950.000)
```

```
-----------------------------------------------
Observations                    20,778
R2                              0.220
Adjusted R2                     0.220
Residual Std. Error 347,109,663.000 (df = 20771)
F Statistic          978.627*** (df = 6; 20771)
===============================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

Anexo 5. Resultados modelo *random forests*

*Se muestran resultados con las funciones del Anexo 2

- Bogotá

```
rpart(formula = chapi_1, data = chapi_train)
  n= 13451

          CP nsplit rel error    xerror       xstd
1 0.24276019      0 1.0000000 1.0001351 0.01662896
2 0.06058908      1 0.7572398 0.7574052 0.01330288
3 0.02813650      2 0.6966507 0.6970183 0.01281635
4 0.01504773      3 0.6685142 0.6689865 0.01247372
5 0.01360231      4 0.6534665 0.6562842 0.01233866
6 0.01007209      5 0.6398642 0.6404498 0.01233783
7 0.01000000      6 0.6297921 0.6361757 0.01220810

Node number 1: 13451 observations,    complexity param=0.2427602
  mean=1.330578e+09, MSE=8.194878e+17

Node number 2: 5537 observations,    complexity param=0.01360231
  mean=7.973401e+08, MSE=2.343999e+17

Node number 3: 7914 observations,    complexity param=0.06058908
  mean=1.703655e+09, MSE=8.907163e+17

Node number 4: 1922 observations
  mean=5.716591e+08, MSE=1.519596e+17

Node number 5: 3615 observations
  mean=9.173287e+08, MSE=2.367547e+17

Node number 6: 2455 observations,    complexity param=0.01007209
  mean=1.270465e+09, MSE=6.218715e+17

Node number 7: 5459 observations,    complexity param=0.0281365
  mean=1.898468e+09, MSE=8.892774e+17

Node number 12: 2246 observations
  mean=1.205594e+09, MSE=5.286518e+17

Node number 13: 209 observations
  mean=1.967596e+09, MSE=1.092434e+18

Node number 14: 4434 observations,    complexity param=0.01504773
  mean=1.783866e+09, MSE=7.924706e+17

Node number 15: 1025 observations
  mean=2.394218e+09, MSE=1.005467e+18

Node number 28: 2648 observations
  mean=1.625023e+09, MSE=7.025614e+17
```

```
Node number 29: 1786 observations
  mean=2.019373e+09, MSE=8.329016e+17
```

- Medellín

```
          CP nsplit rel error    xerror       xstd
1  0.11579257      0 1.0000000 1.0000234 0.04115690
2  0.07387939      1 0.8842074 0.8844116 0.03641143
3  0.04159905      2 0.8103280 0.8113289 0.03441234
4  0.02189414      3 0.7687290 0.7768021 0.03264601
5  0.01717236      4 0.7468348 0.7625158 0.03165765
6  0.01704833      6 0.7124901 0.7497422 0.03126828
7  0.01670847      7 0.6954418 0.7438241 0.03114595
8  0.01263356      8 0.6787333 0.7067019 0.03010065
9  0.01085074      9 0.6660998 0.6983357 0.02991569
10 0.01000000     10 0.6552490 0.6833262 0.02952630

Node number 1: 21356 observations,    complexity param=0.1157926
  mean=4.057746e+08, MSE=1.543341e+17

Node number 2: 16657 observations,    complexity param=0.02189414
  mean=3.347718e+08, MSE=6.256533e+16

Node number 3: 4699 observations,    complexity param=0.07387939
  mean=6.574652e+08, MSE=3.984169e+17

Node number 4: 16540 observations,    complexity param=0.01670847
  mean=3.29236e+08, MSE=5.081554e+16

Node number 5: 117 observations,    complexity param=0.01704833
  mean=1.117356e+09, MSE=1.106832e+18

Node number 6: 4601 observations,    complexity param=0.04159905
  mean=6.242424e+08, MSE=3.308922e+17

Node number 7: 98 observations
  mean=2.217244e+09, MSE=1.083903e+18

Node number 8: 14466 observations,    complexity param=0.01085074
  mean=3.073875e+08, MSE=3.694424e+16

Node number 9: 2074 observations
  mean=4.816277e+08, MSE=1.210141e+17

Node number 10: 51 observations
  mean=3.289941e+08, MSE=7.864364e+16

Node number 11: 66 observations
  mean=1.726545e+09, MSE=1.049969e+18

Node number 12: 4573 observations,    complexity param=0.01717236
  mean=6.107345e+08, MSE=2.947083e+17

Node number 13: 28 observations
  mean=2.830357e+09, MSE=1.343761e+18

Node number 16: 3128 observations
  mean=2.127243e+08, MSE=1.754785e+16

Node number 17: 11338 observations
  mean=3.335037e+08, MSE=3.914113e+16

Node number 24: 1778 observations
  mean=4.759869e+08, MSE=1.592776e+17

Node number 25: 2795 observations,    complexity param=0.01717236
  mean=6.964524e+08, MSE=3.619629e+17
```

```
Node number 50: 2774 observations,    complexity param=0.01263356
  mean=6.836641e+08, MSE=3.369832e+17

Node number 51: 21 observations
  mean=2.385714e+09, MSE=7.864531e+17

Node number 100: 1906 observations
  mean=6.009844e+08, MSE=2.305417e+17

Node number 101: 868 observations
  mean=8.652166e+08, MSE=5.227409e+17
```

Anexo 6. Resultados modelo *rgboost*

*Se utilizaron funciones del Anexo 3

- Bogotá con variables lineales

```
[1] train-rmse:808279567.963563
[2] train-rmse:406096471.352101
[3] train-rmse:204488351.812953
[4] train-rmse:103560884.765471
[5] train-rmse:53086785.424197
[6] train-rmse:27777806.587678
[7] train-rmse:15437987.100543
[8] train-rmse:9775410.998571
[9] train-rmse:7094363.084718
[10] train-rmse:6106198.745394
                ...
[90] train-rmse:847291.034356
[91] train-rmse:824964.702165
[92] train-rmse:816499.637021
[93] train-rmse:800164.701666
[94] train-rmse:794067.800001
[95] train-rmse:791163.976414
[96] train-rmse:784856.895702
[97] train-rmse:779757.163089
[98] train-rmse:769030.498417
[99] train-rmse:753609.328410

[100] train-rmse:744062.242802
```

- Bogotá con variables cuadráticas

```
[1] train-rmse:808279567.963563
[2] train-rmse:406096471.352101
[3] train-rmse:204488351.812953
[4] train-rmse:103560884.765471
[5] train-rmse:53086785.424197
[6] train-rmse:27777806.587678
[7] train-rmse:15437982.006707
[8] train-rmse:9775430.337547
[9] train-rmse:7094391.521749
[10] train-rmse:6106229.176997
                …
 [90] train-rmse:850412.621522
 [91] train-rmse:843992.682137
 [92] train-rmse:833132.373968
 [93] train-rmse:819242.926100
 [94] train-rmse:809384.488354
 [95] train-rmse:792175.665884
 [96] train-rmse:781353.223583
 [97] train-rmse:773620.178781
 [98] train-rmse:768558.363623
 [99] train-rmse:762046.212612
```

```
[100] train-rmse:754732.843842
```

- Medellín con variables lineales

```
[1] train-rmse:285934062.604647
[2] train-rmse:145167592.206906
[3] train-rmse:74237703.864875
[4] train-rmse:38741955.684560
[5] train-rmse:20835244.586817
[6] train-rmse:11897190.180323
[7] train-rmse:7624251.886975
[8] train-rmse:5188370.118611
[9] train-rmse:3899379.330609
[10] train-rmse:3391613.616659
             …
[90] train-rmse:268793.057613
[91] train-rmse:266941.952401
[92] train-rmse:259612.557570
[93] train-rmse:254560.529884
[94] train-rmse:249814.767971
[95] train-rmse:247996.748536
[96] train-rmse:245595.604836
[97] train-rmse:238900.618382
[98] train-rmse:234494.415851
[99] train-rmse:231816.343644
```

```
[100] train-rmse:228834.798396
```

- Medellín con variables cuadráticas

```
[1] train-rmse:285934062.604647
[2] train-rmse:145167592.206906
[3] train-rmse:74237703.864875
[4] train-rmse:38741955.684560
[5] train-rmse:20835244.586817
[6] train-rmse:11897190.180323
[7] train-rmse:7624251.886975
[8] train-rmse:5188370.118611
[9] train-rmse:3899379.330609
[10] train-rmse:3391613.616659
               ...
[90] train-rmse:271773.685040
[91] train-rmse:268357.935748
[92] train-rmse:263797.534826
[93] train-rmse:257132.449216
[94] train-rmse:251942.808618
[95] train-rmse:243669.602275
[96] train-rmse:239561.780599
[97] train-rmse:234525.017628
[98] train-rmse:232744.133850
[99] train-rmse:230493.085039
```

```
[100] train-rmse:227942.050877
```

Anexo 7. Resultados modelo *superlearners* (elegido)

- Bogotá con variables lineales

	Risk	Coef
SL.lm_All	5.317369e+17	0.01583833

SL.rpart_All	5.213720e+17	0.00000000
SL.xgboost_All	3.509389e+17	0.98416167

- Bogotá con variables cuadráticas

	Risk	Coef
SL.lm_All	5.358729e+17	0.01166931
SL.rpart_All	5.214751e+17	0.00000000
SL.xgboost_All	3.523261e+17	0.98833069

- Medellín con variables lineales

	Risk	Coef
SL.lm_All	1.218113e+17	0.0260250
SL.rpart_All	1.029684e+17	0.1004373
SL.xgboost_All	7.596881e+16	0.8735377

- Medellín con variables cuadráticas

	Risk	Coef
SL.lm_All	1.213429e+17	0.04383829
SL.rpart_All	1.032308e+17	0.05110621
SL.xgboost_All	7.384755e+16	0.90505550

Referencias

Chattopadhyay, M., & Mitra, S. K. (2019). Do airbnb host listing attributes influence room pricing homogenously? International Journal of Hospitality Management, 81, 54-64.

Chica-Olmo, J., González-Morales, J. G., & Zafra-Gómez, J. L. (2020). Effects of location on Airbnb apartment pricing in Málaga. Tourism Management, 77, 103981. Deboosere, R., Kerrigan, D. J., c analysis of Airbnb listing prices and revenue. Regional Studies, Regional Science, 6(1), 143-156.

Deboosere, R., Kerrigan, D. J., Wachsmuth, D., & El-Geneidy, A. (2019). Location, location and professionalization: A multilevel hedonic analysis of Airbnb listing prices and revenue. Regional Studies, Regional Science, 6(1), 143-156.

Dudas, G., Kovalcsik, T., Vida, G., Boros, L., & Nagy, G. (2020). Price determinants of Airbnb listing prices in Lake Balaton Touristic Region, Hungary. European Journal of Tourism Research, 24, UNSP 2410.

Garzón Caro, J. M., & Cardozo Perdomo, D. (2019). Cerros orientales de Bogotá: una aplicación hedónica al precio de la vivienda para la localidad de Chapinero para el año 2018.