

hogar

Karina De Sousa

13 de marzo de 2016

Objetivo:

El amigo de un amigo tiene pensado estudiar en Roma durante 2 años a partir de algún momento. En vista de esto, ha decidido usar la plataforma sturent para encontrar un hogar durante sus estudios al ser la única compañía certificada para estudiantes en Roma. Nuestro amigo, al tener poco capital disponible, necesita escoger inteligentemente el hogar en base a una lista de disponibles provista por sturent. Para ello, necesita de su ayuda tratando de predecir el precio de los apartamentos en base a los apartamentos disponibles y escoger, de esta manera, el mejor entre ellos.

Solución:

1. Cargamos los paquetes necesarios,

```
## [1] "xlsx"      "xlsxjars"  "rJava"     "stats"     "graphics"
## [6] "grDevices" "utils"     "datasets"  "methods"   "base"
```

2. Cargamos los datos de los hogares disponibles

PreProcesamiento de los datos

3. Eliminamos las columnas `foto` y `piso`, porque no serán usadas.
4. La columna `direccion` se transformara en una nueva llamada `tiempo`, la cual representa el tiempo en segundo que tarda la persona en ir de la universidad a los distintos apartamentos.

Para esto, se completo la función `parse_data()` de la siguiente forma,

`parse_data()` recibe un `json` y retorna el mismo transformado a un `data.frame`

Luego, se eliminan los saltos de línea y posibles errores que tengan los datos de la columna `direccion` y se usa el API de Google para calcular la columna `tiempo`,

```
## [1] "distrito"    "direccion"   "tipo"        "descripcion"
## [5] "habitaciones" "precio"     "notas"       "tiempo"
```

5. Ahora, convertimos la columna `tipo`, asignando 1 a los apartamentos, 2 a los mini apartamentos y 3 a los estudios.
6. Con la columna `notas` determinamos si el hogar está disponibles para hombres (1), mujeres (2) o ambos (3).
7. Creamos una nueva columna llamada `servicios` con los siguientes valores,

- 1 - Spese Escluse (No incluye cargos por servicios)
- 2 - Tutto Incluso (Todo incluido)
- 3 - Cualquier otro servicio

Creamos las columnas `calefaccion` y `condominio` y `cocina`, las cuales nos dan más información sobre el hogar y obtenemos estos datos de la columna `precio`.

En la columna `precio` dejamos únicamente el valor del hogar.

7. Convertimos la columna `habitaciones` a datos numéricos,

- 1 - Una sola habitación.
- 2 - Dos habitaciones.
- 3 - Tres habitaciones.
- 4 - Cuatro habitaciones.
- 5 - Apartamento completo.
- 6 - Miniapartamento.
- 7 - Estudio.
- 8 - Una cama.
- 9 - Una habitación doble.
- 10 - Dos habitaciones dobles.

8. Creamos las columnas `entrada`, `habitacion` y `cocina`, las cuales nos dan más información sobre el hogar y obtenemos estos datos de la columna `descripcion`.

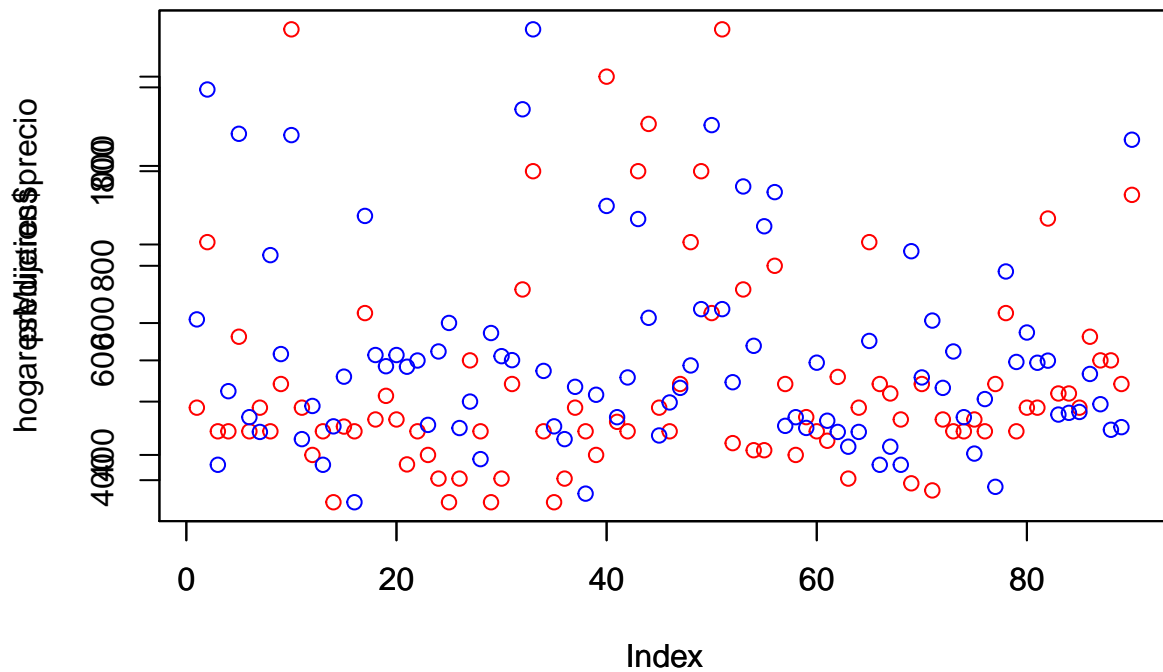
9. Eliminamos las columnas que no son necesarias y así obtenemos una vista minable de los hogares,

```
##      tipo habitaciones precio sexo tiempo servicios calefaccion condominio
## 1      1              1    500    2   13860          1            0          0
## 2      1              1    400    1    360          2            0          0
## 3      1              5    850    2    360          1            0          0
## 4      1              1    450    2    360          3            1          1
## 5      1              2    450    2    360          0            0          0
## 6      1              5    650    3    360          1            0          0
##      entrada habitacion cocina
## 1          1            0      1
## 2          1            0      1
## 3          1            1      0
## 4          1            0      1
## 5          1            0      1
## 6          1            1      1
```

10. Usamos la función `lm` de `stats` para hallar el modelo de regresión lineal que mejor se ajuste a nuestros datos. Separándolos en hogares para mujeres y hogares para hombres.

Modelo para mujeres

11. Seleccionamos los hogares para mujeres y vemos como queda la predicción usando `lm`, donde los puntos azules son los aproximados y los rojos los reales.



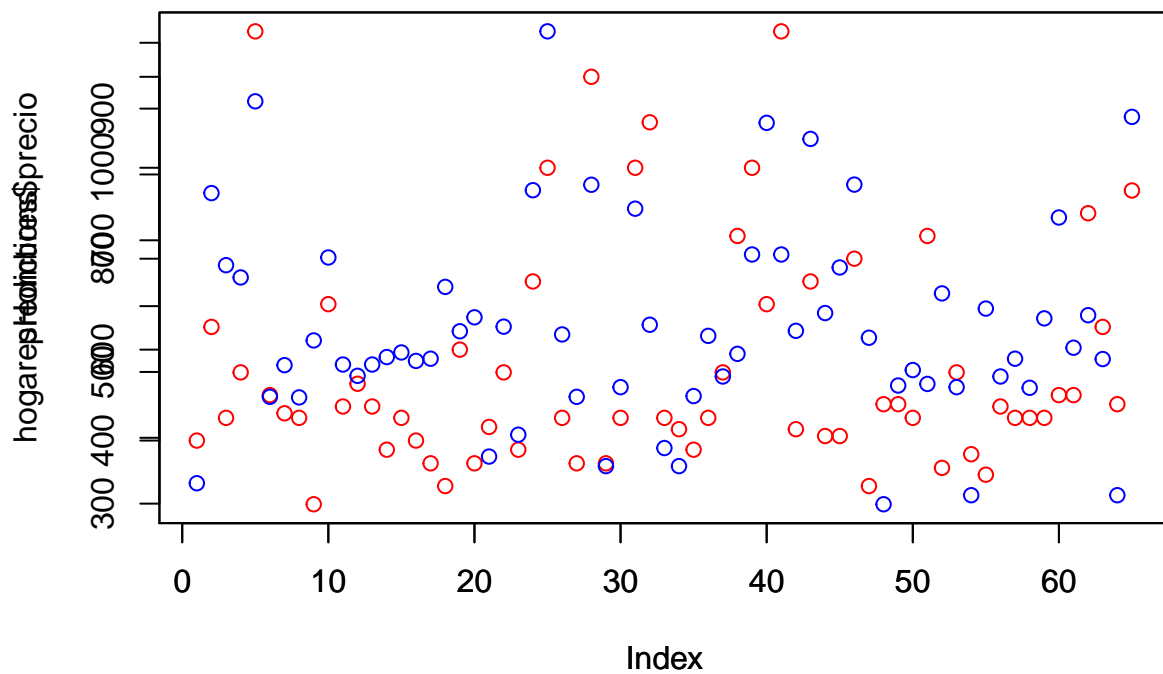
```
## [1] 125.1691
```

Existe una diferencia promedio de 125.16 entre el precio real y el aproximado.

Ahora, veamos que tan buena es la aproximación para los hombres.

Modelo para hombres

12. Seleccionamos los hogares para hombres y vemos como queda la predicción usando `lm`, donde los puntos azules son los aproximados y los rojos los reales.



```
## [1] 143.703
```

13.

Escoja, usando el modelo del primer punto, el mejor apartamento disponible.