

75.06/95.58 Organización de Datos

Finger 2: evolución del precio por m2 en los departamentos de CABA

Puntos: 3,2,1,1,1,1,1

Tiempo Estimado: 2hs

Utilizando Apache Spark se desea obtener:

1. Evolución del precio usd por m2 promedio de las ventas de departamentos en CABA para los últimos 6 meses. Cuál fue el precio usd por m2 promedio (promedio de la columna price_usd_per_m2) para las propiedades publicadas en Febrero 2017, cuál fue para Marzo 2017, ..., Julio 2017.
2. Para cada mes obtener cuáles fueron las 10 propiedades con el menor precio por m2 y cuáles fueron las 10 propiedades con el mayor precio por m2.
3. Y para cada uno de los meses indicar cuantas propiedades tienen +/- 20% del promedio de precio por m2.

Mostrar los datos y para los puntos 1 y 3 realizar una visualización de los mismos.

Los datos son los mismos del TP, que pueden descargarse en

<http://www.properati.com.ar/data/>

Utilizar el dataset [properati-AR-2017-08-01-properties-sell-six_months.csv.gz](#)

Para cargar los datos en Spark pueden utilizar la librería pyspark_csv.

Con pyspark_csv pueden cargar facilmente un CSV en un dataframe de spark de la siguiente manera:

```
import pyspark_csv as pycsv
sc.addPyFile('pyspark_csv.py')
plaintext_rdd = sc.textFile('properati.csv')
dataframe = pycsv.csvToDataFrame(sqlCtx, plaintext_rdd)
data = dataframe.rdd
```

El alumno que complete primero estos 2 puntos subiendo sus resultados al grupo de Facebook de la materia obtiene 3 puntos, el segundo 2 puntos y los 5 siguientes 1 punto cada uno.