

# Mini-Project Part II

## MapReduce scripts

Topic: **Housing prices in Mumbai**

Link: <https://www.kaggle.com/sameep98/housing-prices-in-mumbai>

This data contains scraped data and has the following information:

- 1) Prices of houses all over Mumbai along with their location
- 2) Information about house condition (new/resale) and area of the house
- 3) Information about various amenities provided

### ➤ Dataset information

- Contains 6347 rows
- Contains 18 columns
- The various attributes are:
  1. Price (₹)
  2. Area (sq. ft.)
  3. Location
  4. No. of bedrooms
  5. New/Resale
  6. Gymnasium
  7. Lift available
  8. Car Parking
  9. Maintenance Staff
  10. 24x7 Security
  11. Children's Play Area
  12. Clubhouse
  13. Intercom
  14. Landscaped Gardens
  15. Indoor games
  16. Gas connection
  17. Jogging track
  18. Swimming pool

### 1. Find number of houses in a location (numberofloc.py)

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.18363.1198]
(c) 2019 Microsoft Corporation. All rights reserved.

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python numberofloc.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\numberofloc.Pratik Mali.20201208.163040.172391
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\numberofloc.Pratik Mali.20201208.163040.172391\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\numberofloc.Pratik Mali.20201208.163040.172391\output...
"15th Road"      1
"4 Bungalows"   2
"Adaigaon"      1
"Agripada"      1
"Airoli"        28
"Akurli Nagar"  1
"Almeida Park"  2
"Ambernath"     4
"Ambernath East" 15
"Ambernath West" 15
"Ambivali"      1
"Ambivli"       1
"Anand Nagar Thane West" 1
"Andheri East"  83
"Andheri West" 161
"Andheri"       13
"Anjurdive"     3
"Antop Hill"    5
"Asangaon"      1
"Badlapur East" 40
```

### 2. Find number of houses based on no. of bedrooms (numberbybed.py)

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.18363.1198]
(c) 2019 Microsoft Corporation. All rights reserved.

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python numberbybed.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\numberbybed.Pratik Mali.20201208.163655.116691
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\numberbybed.Pratik Mali.20201208.163655.116691\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\numberbybed.Pratik Mali.20201208.163655.116691\output...
"1"      2307
"2"      2601
"3"      1201
"4"      190
"5"      39
"6"       7
"7"       2
Removing temp directory C:\Users\malip\AppData\Local\Temp\numberbybed.Pratik Mali.20201208.163655.116691...
```

### 3. Find number of houses which are new (numberbynew.py)

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.18363.1198]
(c) 2019 Microsoft Corporation. All rights reserved.

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python numberbynew.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\numberbynew.Pratik Mali.20201208.163933.355302
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\numberbynew.Pratik Mali.20201208.163933.355302\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\numberbynew.Pratik Mali.20201208.163933.355302\output...
"0"      4178
Removing temp directory C:\Users\malip\AppData\Local\Temp\numberbynew.Pratik Mali.20201208.163933.355302...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```

#### 4. Find number of houses which for resale (numberbyold.py)

```
C:\Windows\System32\cmd.exe

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python numberbyold.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\numberbyold.Pratik Mali.20201208.164358.749700
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\numberbyold.Pratik Mali.20201208.164358.749700\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\numberbyold.Pratik Mali.20201208.164358.749700\output...
"1"      2169
Removing temp directory C:\Users\malip\AppData\Local\Temp\numberbyold.Pratik Mali.20201208.164358.749700...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```

#### 5. Find number of luxury flats (numberbyliftcarpool.py)

```
C:\Windows\System32\cmd.exe

Microsoft Windows [Version 10.0.18363.1198]
(c) 2019 Microsoft Corporation. All rights reserved.

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python numberbyliftcarpool.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\numberbyliftcarpool.Pratik Mali.20201208.165243.502090
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\numberbyliftcarpool.Pratik Mali.20201208.165243.502090\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\numberbyliftcarpool.Pratik Mali.20201208.165243.502090\out
put...
"Luxury flat"    2023
Removing temp directory C:\Users\malip\AppData\Local\Temp\numberbyliftcarpool.Pratik Mali.20201208.165243.502090...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```

#### 6. Find number of houses in Powai (numberbypowai.py)

```
C:\Windows\System32\cmd.exe

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python numberbypowai.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\numberbypowai.Pratik Mali.20201208.165553.169546
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\numberbypowai.Pratik Mali.20201208.165553.169546\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\numberbypowai.Pratik Mali.20201208.165553.169546\output...
"Powai" 141
Removing temp directory C:\Users\malip\AppData\Local\Temp\numberbypowai.Pratik Mali.20201208.165553.169546...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```

## 7. Find average rate of house per location in ₹/sq. ft. (ratebyloc.py)

```
C:\Windows\System32\cmd.exe

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python ratebyloc.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\ratebyloc.Pratik Mali.20201208.165944.393601
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\ratebyloc.Pratik Mali.20201208.165944.393601\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\ratebyloc.Pratik Mali.20201208.165944.393601\output...
"15th Road"      43750.0
"4 Bungalows"   22592.59259259259
"Adaigaon"      5087.7192982456145
"Agripada"      53648.0686695279
"Airoli"        12087.70233376855
"Akurli Nagar"  25446.426339285714
"Almeida Park"  43437.5
"Ambarnath"     3816.5725645821763
"Ambernath East" 4403.554295412237
"Ambernath West" 3891.979903240361
"Ambivali"      18264.8401826484
"Ambivli"       3125.0
"Anand Nagar Thane West" 10800.0
"Andheri East"  19676.866887950135
"Andheri West"  27957.09764260826
"Andheri"       19767.22181593159
"Anjurdive"     8517.575757575758
"Antop Hill"    20508.791208791212
"Asangaon"      3545.2961672473866
"Badlapur East" 4256.680953773333
"Badlapur West" 3818.016771257185
"Badlapur"      3965.67971340386
```

## 8. Find average of house based on no. of bedrooms (ratebybed.py)

```
C:\Windows\System32\cmd.exe

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python ratebybed.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\ratebybed.Pratik Mali.20201208.170155.349601
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\ratebybed.Pratik Mali.20201208.170155.349601\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\ratebybed.Pratik Mali.20201208.170155.349601\output...
"1" 9928.955712712404
"2" 13593.402838138009
"3" 18203.19825322239
"4" 24518.689594411837
"5" 26404.149926306673
"6" 25622.912187248956
"7" 22142.85714285714
Removing temp directory C:\Users\malip\AppData\Local\Temp\ratebybed.Pratik Mali.20201208.170155.349601...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```

## 9. Find average rate of new house (ratebynew.py)

```
C:\Windows\System32\cmd.exe

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python ratebynew.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\ratebynew.Pratik Mali.20201208.170321.612189
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\ratebynew.Pratik Mali.20201208.170321.612189\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\ratebynew.Pratik Mali.20201208.170321.612189\output...
"Average rate of new"    13168.88025725882
Removing temp directory C:\Users\malip\AppData\Local\Temp\ratebynew.Pratik Mali.20201208.170321.612189...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```

## 10. Find average rate of new house in powai (ratebypowainew.py)

```
C:\Windows\System32\cmd.exe

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python ratebypowainew.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\ratebypowainew.Pratik Mali.20201208.170727.078829
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\ratebypowainew.Pratik Mali.20201208.170727.078829\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\ratebypowainew.Pratik Mali.20201208.170727.078829\output...
"Rate of new house in powai"    19389.877900655567
Removing temp directory C:\Users\malip\AppData\Local\Temp\ratebypowainew.Pratik Mali.20201208.170727.078829...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```

## 11. Find no. of bedrooms in houses in Powai (bedbyloc.py)

```
C:\Windows\System32\cmd.exe

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>python bedbyloc.py Mumbai.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\malip\AppData\Local\Temp\bedbyloc.Pratik Mali.20201208.171024.512266
Running step 1 of 1...
job output is in C:\Users\malip\AppData\Local\Temp\bedbyloc.Pratik Mali.20201208.171024.512266\output
Streaming final output from C:\Users\malip\AppData\Local\Temp\bedbyloc.Pratik Mali.20201208.171024.512266\output...
"1"    28
"2"    80
"3"    29
"4"    4
Removing temp directory C:\Users\malip\AppData\Local\Temp\bedbyloc.Pratik Mali.20201208.171024.512266...

D:\Pratik\Persistent\Training\Intermediate hadoop\dataset\Mumbai real estate>
```