Lab 1 Exercise 1

Calculate home price statistics

# Objective:

The objective of this exercise is to write a mapper and reducer to calculate the minimum, maximum, and mean price of the homes in each neighborhood represented in the data set. The output of your final solution must look like this:

**Blmngtn 159895.0**

**Blmngtn 194870.88**

**Blmngtn 264561.0**

**Blueste 2.14748365E9**

**Blueste 137500.0**

**Blueste 151000.0**

**BrDale 83000.0**

**BrDale 104493.75**

**BrDale 125000.0**

**BrkSide 52000.0**

**BrkSide 124834.055**

**BrkSide 223500.0**

**ClearCr 130000.0**

**ClearCr 212565.42**

**ClearCr 328000.0**

**CollgCr 110000.0**

**CollgCr 197965.77**

**CollgCr 424870.0**

**Crawfor 90350.0**

**Crawfor 210624.72**

**Crawfor 392500.0**

**Edwards 58500.0**

**Edwards 128219.7**

**Edwards 320000.0**

**Gilbert 141000.0**

**Gilbert 192854.5**

**Gilbert 377500.0**

**IDOTRR 34900.0**

**IDOTRR 100123.78**

**IDOTRR 169500.0**

**MeadowV 75000.0**

**MeadowV 98576.47**

**MeadowV 151400.0**

**Mitchel 84500.0**

**Mitchel 156270.12**

**Mitchel 271000.0**

**NAmes 87500.0**

**NAmes 145847.08**

**NAmes 345000.0**

**NPkVill 127500.0**

**NPkVill 142694.44**

**NPkVill 155000.0**

**NWAmes 82500.0**

**NWAmes 189050.06**

**NWAmes 299800.0**

**NoRidge 190000.0**

**NoRidge 335295.3**

**NoRidge 755000.0**

**NridgHt 154000.0**

**NridgHt 316270.62**

**NridgHt 611657.0**

**OldTown 37900.0**

**OldTown 128225.305**

**OldTown 475000.0**

**SWISU 60000.0**

**SWISU 142591.36**

**SWISU 200000.0**

**Sawyer 62383.0**

**Sawyer 136793.14**

**Sawyer 190000.0**

**SawyerW 76000.0**

**SawyerW 186555.8**

**SawyerW 320000.0**

**Somerst 144152.0**

**Somerst 225379.84**

**Somerst 423000.0**

**StoneBr 170000.0**

**StoneBr 310499.0**

**StoneBr 556581.0**

**Timber 137500.0**

**Timber 242247.45**

**Timber 378500.0**

**Veenker 162500.0**

**Veenker 238772.73**

**Veenker 385000.0**

Note that you may include an arbitrary number of digits after the decimal point for the mean value, because only the integer value of the float will be used in the grading of your result.

# Data:

The data is from a Kaggle competition (<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>). Read the description of the data to determine where the home price is and the neighborhood name is. Ignore all other fields in the data set.

# Driver:

Complete code for the driver has been provided. You do not need to modify any code in the driver for your solution, but you should read the driver code to determine what it's doing.

# Mapper:

Partial code for the mapper has been provided. Implement the TODO's in the mapper code provided to perform the map phase.

# Reducer:

Partial code for the reducer has been provided. Implement the TODO's in the reducer code provided to perform the reduce phase.

# Scripts:

Complete code for the rebuild script has been provided. You do not need to modify any code in the rebuild script for your solution, but you should read the script to determine what it's doing. Syntax for the rebuild script is as follows:

$ **./rebuild.sh**

Complete code for the rerun script has been provided. You do not need to modify any code in the rerun script for your solution, but you should read the script to determine what it's doing. Syntax for the rerun script is as follows:

$ **./rerun.sh**

# How to begin:

Follow these steps for beginning work on your exercise:

1. Login to your sandbox as the user01 user.
2. Create a directory called /user/user01/HOUSES
3. Copy the HOUSES.zip file to the /user/user01/HOUSES directory
4. Unzip the HOUSES.zip file and perform your work from there.
5. Read the rebuild.sh and rerun.sh scripts so you know what they are doing.

# Hints:

Consider the hints below when writing your solution to this exercise.

1. You should first implement the mapper in its entirety before implementing any code in the reducer. In order to test the correctness of your mapper, you should run the launcher as follows:

$ **./rerun.sh –Dmapred.reduce.tasks=0**

1. Once your mapper code is complete, make sure that the output from your mapper is what you expect in your reducer.
2. Do not modify any other code except that which has been specified in the TODO lists.
3. You may, but do not need to, use an IDE to implement this code. You can simply use a text editor along with the rebuild and rerun scripts to determine if your code is compiling and executing correctly.
4. The graders will be using the same rebuild.sh and rerun.sh scripts as have been provided to you when they grade your submissions. Make sure your code builds and executes correctly with these scripts before submitting.

# Submission:

Read very carefully the following instructions for how to create and submit your solution to this exercise. If you have any questions at all, please ask them well in advance of your submission. You will lose points if your submission does not conform to these criteria.

1. Create a zip file ***in your sandbox*** which contains the HousesMapper.java and HousesReducer.java code as follows.

$ **cd /user/user01/HOUSES**

$ **zip LNAME\_FNAME\_L1E1.zip ./HousesMapper.java ./HousesReducer.java**

Replace LNAME with your last name, and FNAME with your first name. So, for a student named John Doe:

$ **zip DOE\_JOHN\_L1E1.zip ./HousesMapper.java ./HousesReducer.java**

1. Upload the zip file to the CANVAS system.