CS 504 Project1

Team member:

Rui Pang, Yiyan zhou

Purpose:

Boston is one of the biggest metros in the US. In this project, we want to find the factors which can influence the property value in Boston. To achieve this, we compare the number of hospitals, police stations, and schools in different areas of Boston to find the correlation between these variables and property value. We believe that these variables can show the general living condition of an area.

Dataset Description:

1. Boston Property Value
2. Boston Police Station Location
3. Boston Private School Location
4. Boston Public School Location
5. Boston Hospital Location

Data Transformation:

1. In Publicschool.py, we calculate the number of public school in an area (based on zip code), and project zip code and number of public schools in that area.
2. In PrivateSchool.py, we calculate the number of private school in an area (based on zip code), and project zip code and number of private schools in that area.
3. In PropertyAssessment.py, we calculate the average price per unit of area of different areas in Boston (based on zip code).
4. In police.py, we calculate the number of police station in different areas.
5. In hospital.py, we calculate the number of hospitals in different areas.
6. In total.py, we sum up the number of public school and private school in Map-Reduce model
7. In facilities.py, we merged police station and number of hospitals with corresponding zip code. The output collection shows the number of such public facilities in the area.
8. In total.py, we combined all retrieved data with average property value. This table clearly shows the correlation between property value and community facilities. We believe that the area with more such community facilities(police station, hospital, school) has higher property values.

Running instructions:

run execute.py
If you need to run in trial mode
run execute.py --trial