Real Estate Information Database System

*Li, Qiaoran*

*Yang, Zhaoxiong*

*Zhang, Liang*

Table of Contents

[INTRODUCTION 2](#_Toc508455944)

[DATA SETS 2](#_Toc508455945)

# Introduction

There are several potential factors that affect a family’s house choice, people may have specific checklist that need to be considered when they are looking for a new house. We deliver a Real Estate Database System which provides a comprehensive information about real estate, helping people searching various factors when they’re hunting a house.

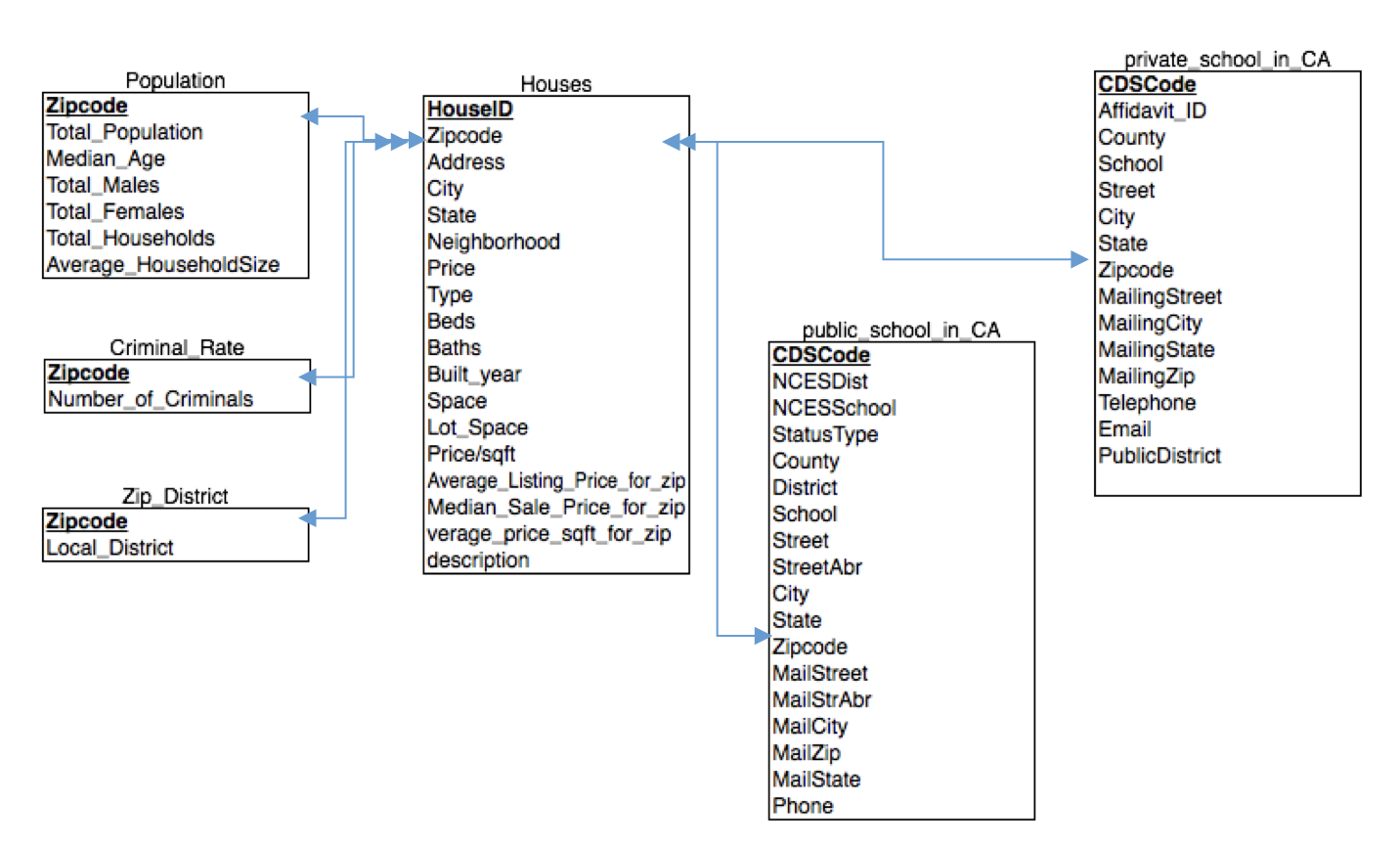
# Solutions

We apply the knowledge we learn in class to prove decision supports taking into consideration of the following factors. (QRcomment: I’m not sure the decision support here is appropriate and sufficient. Maybe I need to think another aspect to connect with the material we learn in class.)

We collect information, which people may take into consideration in home searching process, from several websites by using web crawler (written with Python) and other techniques. Then clean, combine and import them into our database.

# Data Sets and Relational Schema

Our database includes several tables, such as house, crime rate, zip code and neighborhood relation, school, etc. Based on the database we created, we use SQL query to provide diversified information.



*Figure 1: Relational Schema*

As shown in *Figure 1*, there are seven relations in our schema, we connect each other with Zipcode. Since the information about population, criminal rate and school are mainly described in terms of region represented by Zipcode. Data of Population and Criminal Rate is from U.S. Census. And the school data sets of both public and private is from the website in California department of education. Most importantly, the data of houses relation is collected from Trulia.com, using the “web crawler” implemented by Python. We’ll get into detail about it in the Technique Approach section.

The house relation provides the information about available real estate in Los Angeles, including the location, the type of house describing if this house serves single family or multiple family), the number of bed and bath provided, the year of this house built and so on.

Public and private school in CA relation filled in the information as their names, the detail of the school. By connecting the school relation with house relation, the users are able to search the school information based on their potential house location.

The relations such as Population and Criminal provides user with regional information. The population relation gives the particular group of population in a region based on zipcode. Also, the criminal rate relation, XXXXXXX (QRcomments: maybe talk about how we calculate the number of rate, and the integration method.)

# Technique Approach:

We apply the knowledge we learn in class to prove decision supports taking into consideration of the following

1. web crawler: we wrote web crawler using Python. This program is used to gather detailed house information from websites.
2. SQL database

# Use Cases (Query Sample):

--Someone wants to find houses with price <= $500000 and population <= 10000

select h.HouseID, h.Zipcode,p.Total\_Population, h.Address,h.Neighborhood,h.Price

from houses h, population p

where h.HouseID in

(SELECT h1.HouseID FROM Houses h1 WHERE h1.Price <= 500000 and h1.Price > 0)

and h.Zipcode = p.Zipcode

and p.Zipcode in

(select p1.Zipcode from population p1 where p1.Total\_Population <= 10000)

order by p.Total\_Population asc;



-- Someone wants to find houses that near the private high school "Harvard-westlake School"

(select Zipcode, HouseID, Address, Price

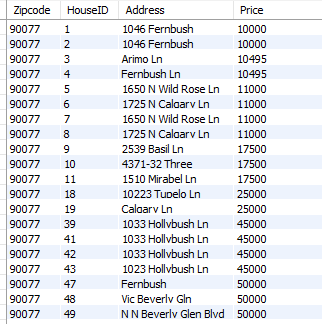
from houses

where Price > 0 and Zipcode in

(select Zipcode

from private\_school\_in\_ca

where School like '%Harvard%'));



-- Someone wants to find houses that near the public school "Sunrise Placement Community Da" and price lower than $500000

select Zipcode, HouseID, Address, Price

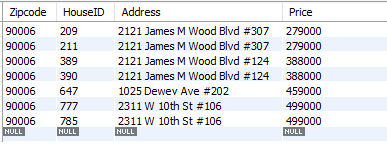
from houses

where Price > 0 and Price <= 500000 and Zipcode in

(select Zipcode

from public\_school\_in\_ca

where School like '%Sunrise Placement Community Da%');

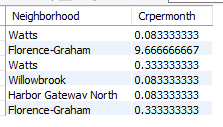


-- Someone wants to find neighbors whose crime cases per month <=10

select distinct(h.Neighborhood), c.Crpermonth

from houses h, crime\_zip c

where c.﻿Zipcode = h.Zipcode and c.Crpermonth<=10 and h.Neighborhood <> ''



*Most of the datasets are for Los Angeles*

*crime-rate: ﻿Zipcode, number of crimes*

*houses: HouseID, Zipcode, Address, City, State, Neighborhood, Price, Type, Beds, Baths, Build year, Space, Lot\_space, Price/sqft, Average\_Listing\_Price\_for\_zip, Median\_Sale\_Price\_for\_zip, Average\_price\_sqft\_for\_zip, description*

*population: Zipcode, Total\_population, Median\_Age, Total\_males, Total\_females, Total\_households, Average\_householdsize*

*private\_school: ID, Zipcode, Address, School\_type, Enrollment, FullTime\_teacher, etc.*

*public\_school: ID, Zipcode, Address, School\_type, Enrollment, FullTime\_teacher, etc.*

*zip-district: relations of zip code and district*

*zip\_lat\_long: relations of zip code and Latitude and longitude coordinates*

(QRcomments: 1. Since we’ve already have the figure 1, we propose to delete the gray part. 2.The relational schema is not the same with our actual data base schema, since there are 64 columns in private school and population school, I omit some column, just keep the essential ones.)