**Create Table:**

CREATE TABLE if not exists CITY\_ADDRESSES

(

Serial\_No double,PIN STRING,PIND STRING,HSE\_NBR double,HSE\_FRAC\_NBR STRING,

HSE\_DIR\_CD STRING,STR\_NM STRING,STR\_SFX\_CD STRING,STR\_SFX\_DIR\_CD STRING,

UNIT\_RANGE STRING,ZIP\_CD double,LAT double,LON double,X\_COORD\_NBR double,

Y\_COORD\_NBR double,ASGN\_STTS\_IND String,ENG\_DIST STRING,CNCL\_DIST STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

TBLPROPERTIES("skip.header.line.count"="1");

**Loading table:**

LOAD DATA INPATH '\HdiSamples\Addresses\_in\_the\_City\_of\_Los\_Angeles.csv'OVERWRITE INTO TABLE CITY\_ADDRESSES;

**Query to calculate the number of houses that are vacant and occupied in LA:**

Select ASGN\_STTS\_IND,COUNT(ASGN\_STTS\_IND) AS cnt FROM CITY\_ADDRESSES GROUP BY ASGN\_STTS\_IND

ORDER BY cnt DESC;

**Query to calculate the number of houses in all the four direction of LA:**

Select HSE\_DIR\_CD,COUNT(HSE\_DIR\_CD) AS cnt FROM CITY\_ADDRESSES GROUP BY HSE\_DIR\_CD

ORDER BY cnt DESC;

**Query to find the top 15 street name that has more number of houses vacant:**

Select STR\_NM,COUNT(STR\_NM) as cnt FROM CITY\_ADDRESSES WHERE ASGN\_STTS\_IND='U' GROUP BY STR\_NM Order BY cnt DESC;

**Create another table:**

Create table if not exists location\_dis1(LAT double,LON double,DIR STRING,ASGN\_STTS\_IND STRING,distance double);

**Override into table:**

INSERT OVERWRITE TABLE location\_dis1 Select LAT,LON,HSE\_DIR\_CD,ASGN\_STTS\_IND,

2 \* asin(

sqrt(

cos(radians(34.0667))\*

cos(radians(LAT)) \*

pow(sin(radians(-118.1678 - LON)/2),2)

+

pow(sin(radians(34.0667 - LAT)/2),2)

)) \*3956

**Query to point all the house that are registered in LA:**

SELECT location\_dis1.lat, location\_dis1.lon, location\_dis1.dir, location\_dis1.asgn\_stts\_ind, location\_dis1.distance

FROM HIVE.default.location\_dis1 location\_dis1;

**Query to point the house that are registered in and around csula in a radius of 5 miles in LA:**

SELECT location\_dis1.lat, location\_dis1.lon, location\_dis1.dir, location\_dis1.asgn\_stts\_ind, location\_dis1.distance

FROM HIVE.default.location\_dis1 location\_dis1 where location\_dis1.distance between 0 and 5;

Query to point the houses that are in LA which are occupied.

SELECT location\_dis1.lat, location\_dis1.lon, location\_dis1.dir, location\_dis1.asgn\_stts\_ind, location\_dis1.distance

FROM HIVE.default.location\_dis1 location\_dis1 where location\_dis1.asgn\_stts\_ind='A';

Query to point the houses that are in LA which are unoccupied:

SELECT location\_dis1.lat, location\_dis1.lon, location\_dis1.dir, location\_dis1.asgn\_stts\_ind, location\_dis1.distance

FROM HIVE.default.location\_dis1 location\_dis1 where location\_dis1.asgn\_stts\_ind='U';

Query to find the total number of houses that are available in a particular house from CSULA which are vacant.

select location\_dis1.distance as score, count(\*) as occurences

from (

select case

when distance between 0 and 5 then ' 0- 5'

when distance between 5 and 10 then '5-10'

when distance between 10 and 15 then '10-15'

when distance between 15 and 20 then '15-20'

when distance between 20 and 25 then '20-25'

when distance between 25 and 30 then '25-30'

when distance between 30 and 35 then '30-35'

when distance between 35 and 40 then '35-40'

when distance between 40 and 50 then '40-50'

else '>50'

end as distance

from location\_dis1 where ASGN\_STTS\_IND='U' ) location\_dis1

group by location\_dis1.distance

order by occurences;

Dataset Link:

<http://catalog.data.gov/dataset/addresses-in-the-city-of-los-angeles>