**Quering**

**First cleaning**

Removing duplicates , converting date column to appropriate type

create table my-first-327813.house\_sales.sales as(  
with table\_1 as (  
SELECT \*,  
CAST(date as string ) as date\_str,  
  
FROM `my-first-327813.house\_sales.sales\_main`   
WHERE   
id IS NOT NULL AND  
date IS NOT NULL AND  
price IS NOT NULL AND  
bedrooms IS NOT NULL AND   
bathrooms IS NOT NULL AND  
sqft\_living IS NOT NULL AND  
sqft\_lot IS NOT NULL AND   
floors IS NOT NULL AND  
waterfront IS NOT NULL AND  
view IS NOT NULL AND   
condition IS NOT NULL AND   
grade IS NOT NULL AND  
sqft\_above IS NOT NULL AND   
sqft\_basement IS NOT NULL AND   
yr\_built IS NOT NULL AND   
yr\_renovated IS NOT NULL AND  
zipcode IS NOT NULL AND  
lat IS NOT NULL AND  
long IS NOT NULL AND   
sqft\_living15 IS NOT NULL AND   
sqft\_lot15 IS NOT NULL   
  
),  
table\_2 as (  
select \*,  
substr( date\_str ,1,4 ) as year,  
substr(date\_str,5,2) as month,  
substr(date\_str,7,2) as day,  
from table\_1 )  
  
select cast(date\_str\_1 as date ) as sales\_date,\*  
from(select \*,  
concat (year,"-",month,"-",day) as date\_str\_1   
from table\_2 )  
);

Removing columns

alter table my-first-327813.house\_sales.sales drop column year;  
alter table my-first-327813.house\_sales.sales drop column month;  
alter table my-first-327813.house\_sales.sales drop column day;  
alter table my-first-327813.house\_sales.sales drop column date ;  
alter table my-first-327813.house\_sales.sales drop column date\_str;  
alter table my-first-327813.house\_sales.sales drop column date\_str\_1;

**Main query** :

* Removing Null values
* Grouping housing by similar specifications
* Calculating of difference in renovated and not renovated housing

WITH table\_1 AS (

select

price,

bedrooms,

#creatign categories for bathrooms

CASE

WHEN bathrooms BETWEEN 0 AND 1.4 THEN 1

WHEN bathrooms BETWEEN 1.5 AND 2.4 THEN 2

WHEN bathrooms BETWEEN 2.5 AND 3.4 THEN 3

WHEN bathrooms BETWEEN 3.5 AND 4.4 THEN 4

        WHEN bathrooms BETWEEN 4.5 AND 5.4 THEN 5

        WHEN bathrooms BETWEEN 5.5 AND 6.4 THEN 6

        WHEN bathrooms BETWEEN 6.5 AND 7.4 THEN 7

        WHEN bathrooms BETWEEN 7.5 AND 8.5 THEN 8

        END AS bathrooms,

# creating groups for sqft\_living (main categories)

    CASE

        WHEN sqft\_living BETWEEN 0 AND 500 THEN "0-500"

        WHEN sqft\_living BETWEEN 501 AND 1000 THEN "500-1000"

        WHEN sqft\_living BETWEEN 1001 AND 1500 THEN "1000-1500"

        WHEN sqft\_living BETWEEN 1501 AND 2000 THEN "1500-2000"

        WHEN sqft\_living BETWEEN 2001 AND 5000 THEN "2000-5000"

        WHEN sqft\_living BETWEEN 5001 AND 10000 THEN "5000-10000"

        END AS sqft\_living,

    CASE

        WHEN sqft\_lot BETWEEN 0 AND 1000 THEN "0-1000"

        WHEN sqft\_lot BETWEEN 1001 AND 2000 THEN "1000-2000"

        WHEN sqft\_lot BETWEEN 2001 AND 5000 THEN "2000-5000"

        WHEN sqft\_lot BETWEEN 5001 AND 10000 THEN "5000-10000"

        WHEN sqft\_lot BETWEEN 10001 AND 20000 THEN "10000-20000"

        WHEN sqft\_lot BETWEEN 20001 AND 50000 THEN "20000-50000"

        When sqft\_lot BETWEEN 50001 AND 150000 THEN "50000-150000"

        When sqft\_lot BETWEEN 150001 AND 1000000 THEN "150000-1000000"

        END AS sqft\_lot,

                floors,

                condition,

                    grade,

                    zipcode,

                    yr\_renovated

from `my-first-327813.house\_sales.housing\_clean`),

renov as (

    SELECT AVG(price ) as avg\_price,

            bedrooms,

            bathrooms,

            sqft\_living,

            sqft\_lot,

            floors,

            condition,

            grade,

            zipcode

    FROM table\_1

    WHERE yr\_renovated > 0

    GROUP BY bedrooms,

            bathrooms,

            sqft\_living,

            sqft\_lot,

            floors,

            condition,

            grade,

            zipcode

),

not\_renov as (

    SELECT AVG(price ) as avg\_price,

            bedrooms,

            bathrooms,

            sqft\_living,

            sqft\_lot,

            floors,

            condition,

            grade,

            zipcode

    FROM table\_1

    WHERE yr\_renovated = 0

    GROUP BY

            bedrooms,

            bathrooms,

            sqft\_living,

            sqft\_lot,

            floors,

            condition,

            grade,

            zipcode

),

table\_4 as (

SELECT

    f.bedrooms,

    f.bathrooms,

    f.sqft\_living,

    f.sqft\_lot,

    f.floors,

    f.condition,

    f.grade,

    f.zipcode,

    avg(f.avg\_price) as avg\_price\_renov,

    avg(s.avg\_price) as avg\_price\_not\_renov,

    avg((f.avg\_price - s.avg\_price) / s.avg\_price ) as avg\_diff

FROM renov  as f

JOIN not\_renov as s on f.bedrooms = s.bedrooms and

            f.bathrooms =s.bathrooms and

            f.sqft\_living = s.sqft\_living and

            f.sqft\_lot = s.sqft\_lot and

            f.floors = s.floors and

            f.condition = s.condition and

            f.grade = s.grade and

            f.zipcode = f.zipcode

GROUP BY

    f.bedrooms,

    f.bathrooms,

    f.sqft\_living,

    f.sqft\_lot,

    f.floors,

    f.condition,

    f.grade,

    f.zipcode)

select \*

from table\_4

**Sales over the year (05/04/2014 -05/10/2015) grouped by week:**

Select

    date\_trunc(sales\_date, week) as week,

    count(id) as houses\_sold

From `my-first-327813.house\_sales.housing\_clean`

Group by week

ORDER BY week asc

**Sales over week:**

with table\_1 as (

SELECT

    FORMAT\_DATE('%A', sales\_date) AS weekday,

    round(avg(count\_id )) as count\_id,

    round(avg(avg\_pice)) as avg\_price,

FROM (

SELECT

    sales\_date,

    count(id) as count\_id,

    round(avg(price)) as avg\_pice

FROM

    `my-first-327813.house\_sales.housing\_clean`

GROUP BY

    sales\_date)

GROUP BY

    weekday),

table\_2 as (

SELECT  \*,

    CASE

        WHEN weekday = "Monday" THEN 1

        WHEN weekday = "Tuesday" THEN 2

        WHEN weekday = "Wednesday" THEN 3

        WHEN weekday = "Thursday" THEN 4

        WHEN weekday = "Friday" THEN 5

        WHEN weekday = "Saturday" THEN 6

        WHEN weekday = "Sunday" THEN 7

    END as sequense

FROM

    table\_1

ORDER BY

    sequense

)

SELECT

    weekday,

    count\_id,

    avg\_price

FROM

    table\_2

**Avg\_diff to coordinates on map**

with table\_1 as (

    select zipcode,

        avg(avg\_diff) as avg\_diff

    from `my-first-327813.house\_sales.diff\_grouped`

    group by zipcode

)

select  d.zipcode,

        lat,

        long,

        d.avg\_diff

from table\_1 as d

Join my-first-327813.house\_sales.housing\_clean as c on d.zipcode = c.zipcode

**AVG\_Diff to yer\_renovation , yr\_built**

with table\_1 as (

SELECT zipcode ,

        avg(yr\_renovated) as yr\_renovated

FROM `my-first-327813.house\_sales.housing\_clean`

WHERE yr\_renovated > 0

Group by zipcode

)

select

        y.zipcode,

        round(avg(avg\_diff),2) as avg\_diff,

        round(avg(y.yr\_renovated)) as avg\_renovated ,

        round(avg(c.yr\_built)) as avg\_built

FROM `my-first-327813.house\_sales.renov\_diff` as d

JOIN `my-first-327813.house\_sales.housing\_clean`  as c ON d.zipcode = c.zipcode

JOIN table\_1 as y on y.zipcode = d.zipcode

GROUP BY y.zipcode

**Renovation effect**

With table\_1 as (

SELECT

    zipcode,

    round(avg(avg\_diff),3) as avg\_diff

FROM `my-first-327813.house\_sales.renov\_diff`

GROUP BY zipcode

)

SELECT

    o.zipcode,

    lat,

    long,

    avg\_diff,

    CASE

        WHEN o.avg\_diff < 0 THEN -1

        WHEN o.avg\_diff >= 0 THEN 1

        END AS renov\_effect

FROM table\_1 as o

JOIN `my-first-327813.house\_sales.housing\_clean` as c on o.zipcode = c.zipcode