***1)What are the most and least expensive houses in the data set?***

SELECT MAX(price) as max\_price, MIN(price)as min\_price FROM home\_data

|  |  |
| --- | --- |
| max\_price | min\_price |
| 7700000 | 75000 |

***2) What is the most expensive zipcode in the data set, defined as highest average sales price?***

select zipcode, avg(price) as average from home\_data GROUP BY zipcode order by average desc limit 1

|  |  |
| --- | --- |
| zipcode | average |
| 98039 | 2160606.6 |

{for the rarest condition when 2 or more zip code have the exact same highest average sales...

SELECT zipcode, AVG(price) AS average

FROM home\_data GROUP BY zipcode

HAVING average IN (

SELECT max(average) FROM (

SELECT zipcode, avg(price) AS average

FROM home\_data

GROUP BY zipcode

ORDER BY average DESC

) AS temp)

}

***3) How many houses were built prior to 1979?***

SELECT COUNT(yr\_built) AS HOUSES\_BULT\_BEFORE\_1979 FROM HOME\_DATA WHERE yr\_built<1979

|  |
| --- |
| houses\_bult\_before\_1979 |
| 11991 |

***4)How many homes were sold with a zipcode defined as being in “Seattle”?***

SELECT count(h.zipcode) as count FROM HOME\_DATA h join wa\_zipcodes w on (h.zipcode = w.zipcode) where(UPPER(w.city) = 'SEATTLE');

|  |
| --- |
| count |
| 8977 |

***5) Output a report showing the most expensive house in each city. Include at a minimum***

***the price, zipcode and city.***

select \* from ( select h.id, h.price, h.zipcode, w.city,

row\_number() over (partition by w.city order by h.price desc) as myNumber

from home\_data h join wa\_zipcodes w on h.zipcode = w.zipcode) as temp

where myNumber = 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| temp.id | temp.price | temp.zipcode | temp.city | temp.mynumber |
| 2025770560 | 930000 | 98092 | Auburn | 1 |
| 9808700762 | 7062500 | 98004 | Bellevue | 1 |
| 5416300240 | 935000 | 98010 | Black Diamond | 1 |
| 1726059134 | 1075000 | 98011 | Bothell | 1 |
| 853600150 | 1680000 | 98014 | Carnation | 1 |
| 2526069092 | 1015000 | 98019 | Duvall | 1 |
| 1020069042 | 858000 | 98022 | Enumclaw | 1 |
| 1524079188 | 1862000 | 98024 | Fall City | 1 |
| 8894200150 | 1275000 | 98023 | Federal Way | 1 |
| 2524069078 | 2700000 | 98027 | Issaquah | 1 |
| 7403200050 | 1600000 | 98028 | Kenmore | 1 |
| 2322059136 | 859000 | 98042 | Kent | 1 |
| 1247600105 | 5110800 | 98033 | Kirkland | 1 |
| 822069112 | 1350000 | 98038 | Maple Valley | 1 |
| 9208900037 | 6885000 | 98039 | Medina | 1 |
| 7558700030 | 5300000 | 98040 | Mercer Island | 1 |
| 8835800350 | 1950000 | 98045 | North Bend | 1 |
| 1225069038 | 2280000 | 98053 | Redmond | 1 |
| 518500480 | 3000000 | 98056 | Renton | 1 |
| 624069108 | 3200000 | 98075 | Sammamish | 1 |
| 6762700020 | 7700000 | 98102 | Seattle | 1 |
| 2624089007 | 1998000 | 98065 | Snoqualmie | 1 |
| 1269200229 | 1379900 | 98070 | Vashon | 1 |
| 1630700380 | 1920000 | 98077 | Woodinville | 1 |

***Bonus Exercise 1:***

***Price per sq. ft. is a common metric for valuing housing. Calculate the avg. price per sq. ft for***

***the houses sold in this data set.***

select AVG(price/sqft\_living) as avg\_price\_per\_sq\_ft from home\_data

|  |
| --- |
| avg\_price\_per\_sq\_ft |
| 264.15659380747553 |

***Bonus Exercise 2:***

***“date” is a reserved keyword in Hive. The sample data has a column named “date”. How can***

***you use this column in a query without an error? Generate a list of all the unique date values***

***represented in this data set.***

SELECT DISTINCT `date` FROM `home\_data`;

372 distinct dates in the dataset. (too big of a list to print.)

Keywords can be used as identifiers by using backticks.  
for example: SELECT `date` FROM `database`.`table`;