**Tables:**

**Data**

* Userid NUMBER
* Itemid NUMBER
* Rating NUMBER
* Timestamp NUMBER

**Movie**

* movieid (theta join with item id) NUMBER
* Title VARCHAR2(255)
* Release\_date VARCHAR2(30)
* Unknown NUMBER
* Action NUMBER
* Adventure NUMBER
* Animation NUMBER
* Childrens NUMBER
* Comedy NUMBER
* Crime NUMBER
* Documentary NUMBER
* Drama NUMBER
* Fantasy NUMBER
* Film\_Noir NUMBER
* Horror NUMBER
* Musical NUMBER
* Mystery NUMBER
* Romance NUMBER
* Sci\_Fi NUMBER
* Thriller NUMBER
* War NUMBER
* Western NUMBER

**Users**

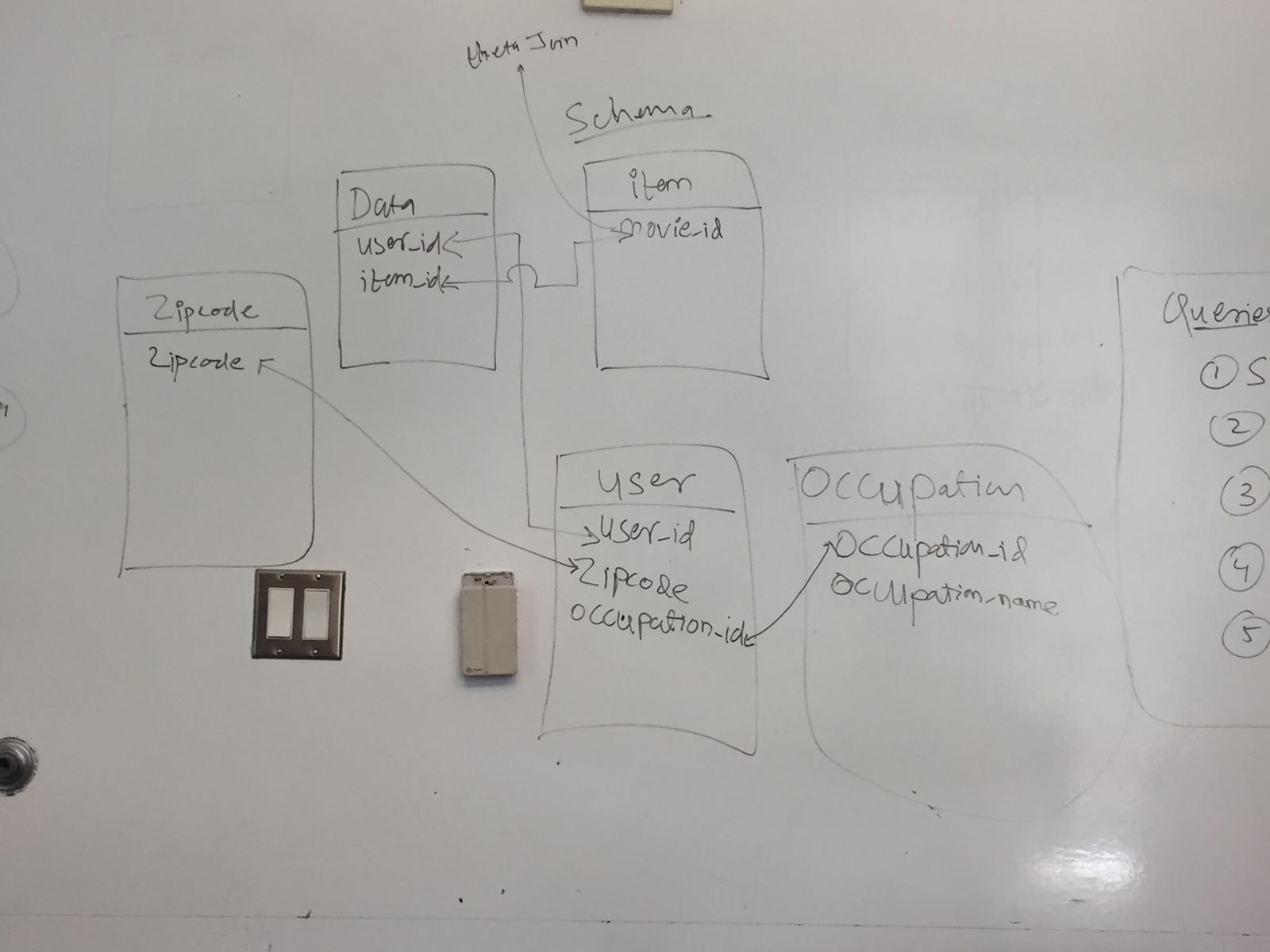
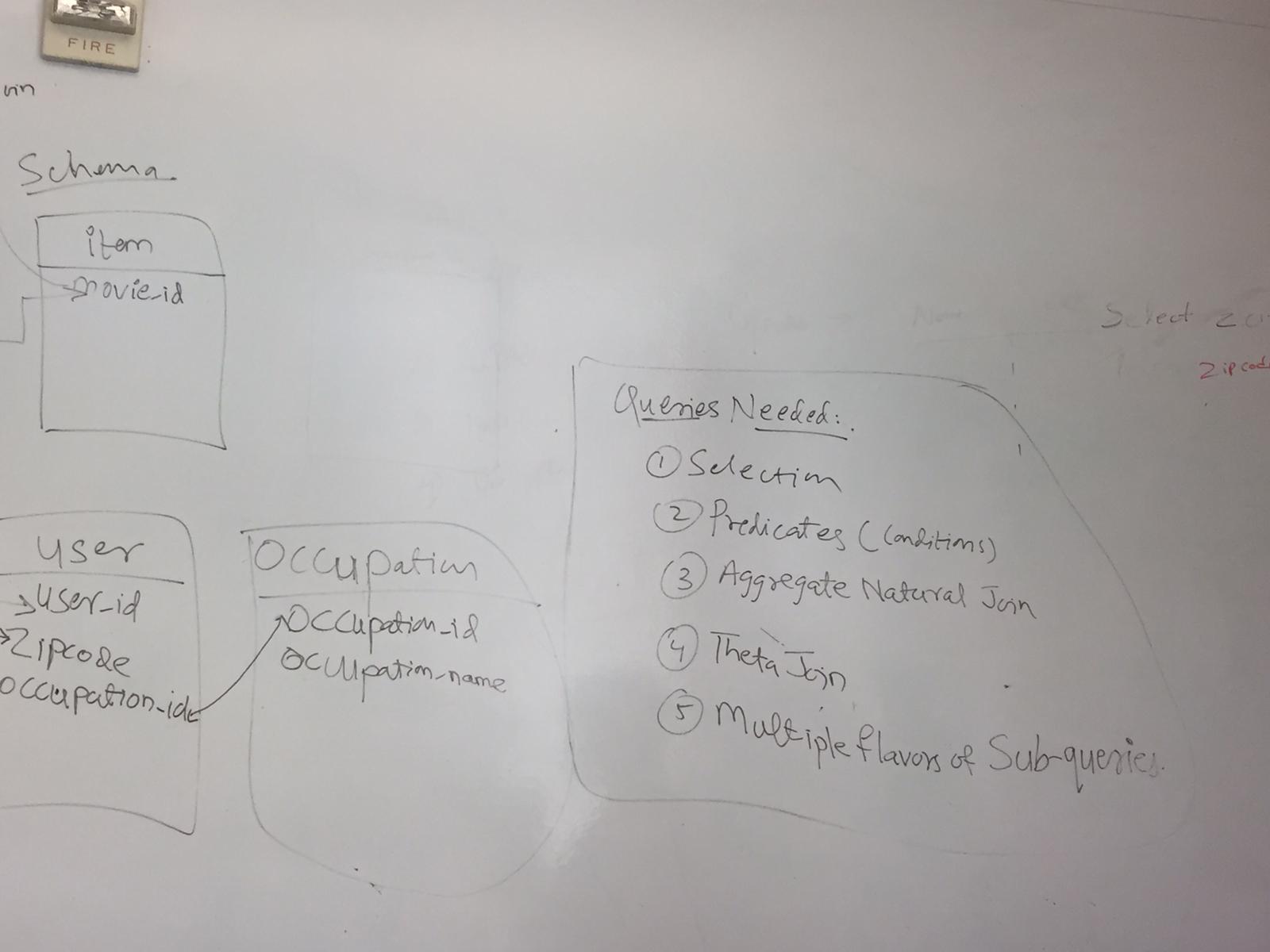
* Userid NUMBER
* Age NUMBER
* Gender VARCHAR2(1)
* Occupationid NUMBER
* Zipcode VARCHAR2(30)

**Occupation**

* Occupationid NUMBER
* Occupation VARCHAR2(30)

**Zipcode**

* Zipcode VARCHAR2(30)
* City VARCHAR2(30)
* State VARCHAR2(30)

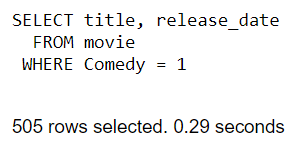


1. Select movie where comedy is 1

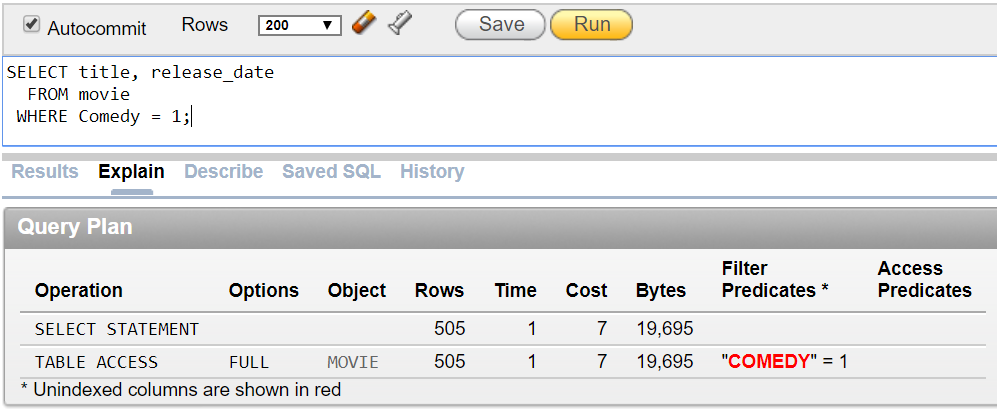
SELECT title, release\_date

FROM movie

WHERE Comedy = 1;​







1. Most popular movie among women

SELECT m.movieid,

m.title

FROM movie m

INNER JOIN (

SELECT d.userid,

d.itemid,

d.rating

FROM data d

INNER JOIN (

SELECT userid,

gender

FROM users

WHERE gender = 'F'

) f

ON d.userid = f.userid

WHERE d.rating = 5

) d

ON m.movieid = d.itemid​



1. Average rating for each gender

SELECT u.gender,

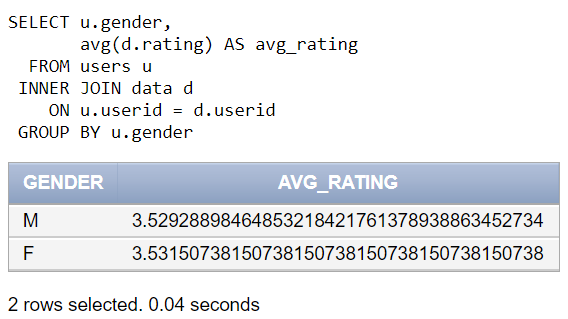
avg(d.rating) AS avg\_rating

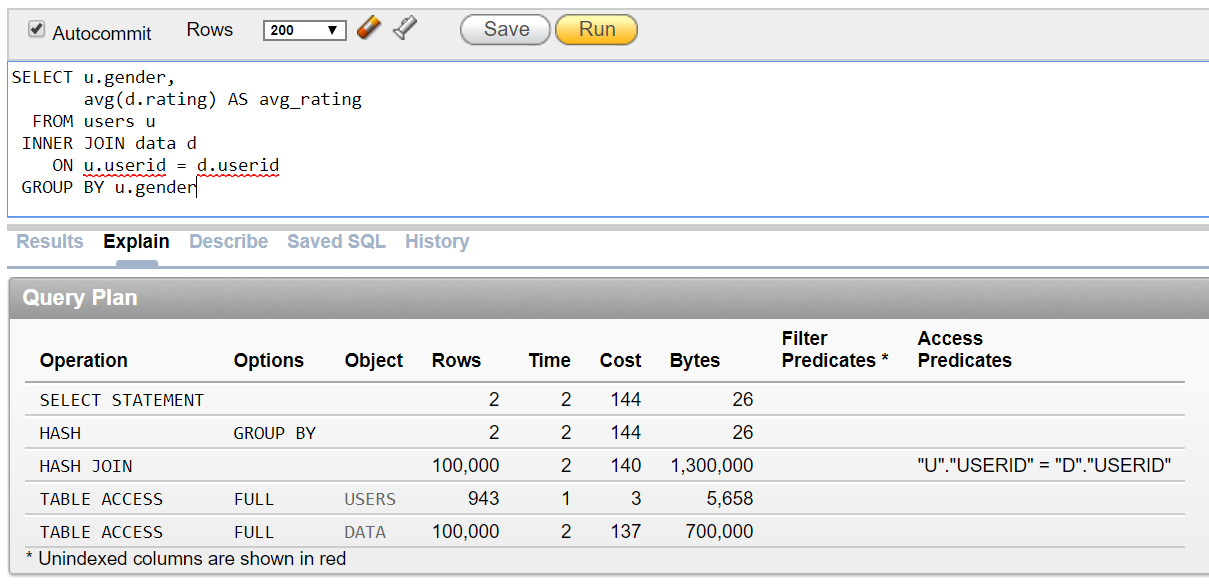
FROM users u

INNER JOIN data d

ON u.userid = d.userid

GROUP BY u.gender;​





1. Rank movies by average rating

SELECT m.title,

avg(d.rating) AS avg\_rating

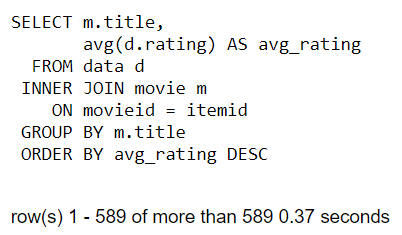
FROM data d

INNER JOIN movie m

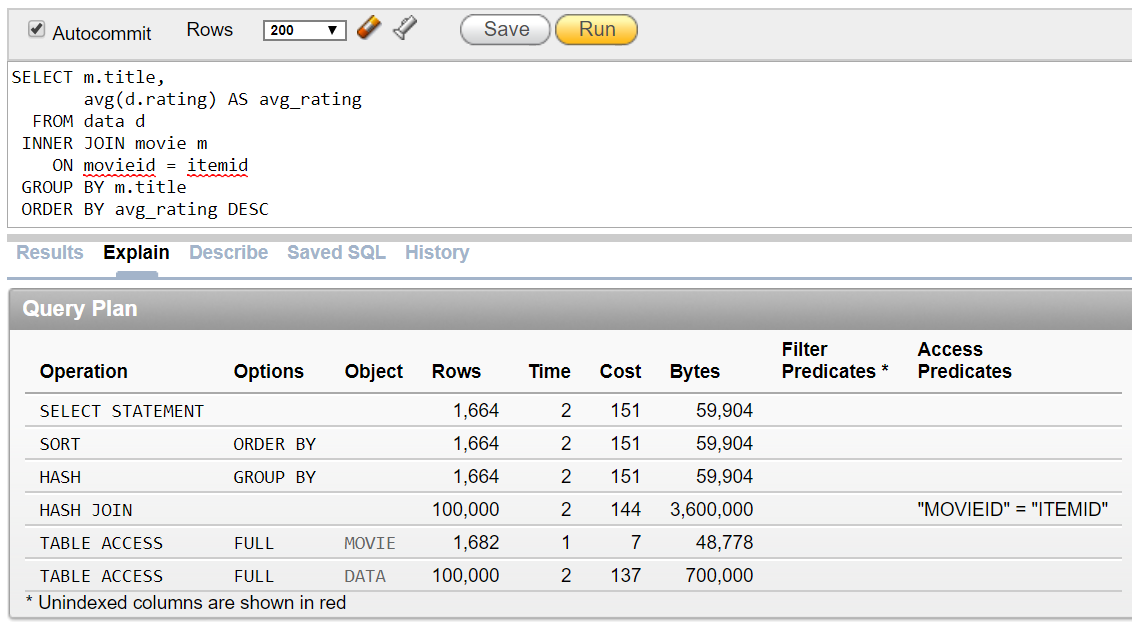
ON movieid = itemid

GROUP BY m.title

ORDER BY avg\_rating DESC;







1. Select movies watched where the users live in Santa Clara and are not an engineer

SELECT DISTINCT m.title

FROM movie m

WHERE m.movieid IN (

SELECT d.itemid

FROM data d

WHERE d.userid IN (

SELECT u.userid

FROM users u

WHERE u.occupationid IN (

SELECT o.occupationid

FROM occupation o

WHERE o.occupation != 'engineer'

)

AND u.zipcode IN (

--zipcodes of Santa Clara

SELECT z.zipcode

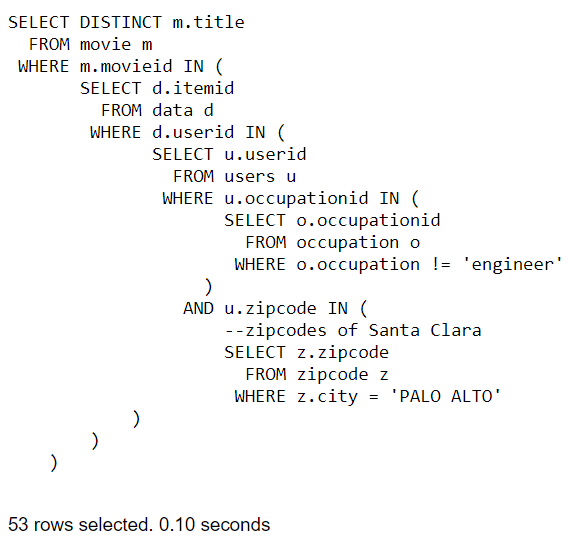
FROM zipcode z

WHERE z.city = 'PALO ALTO'

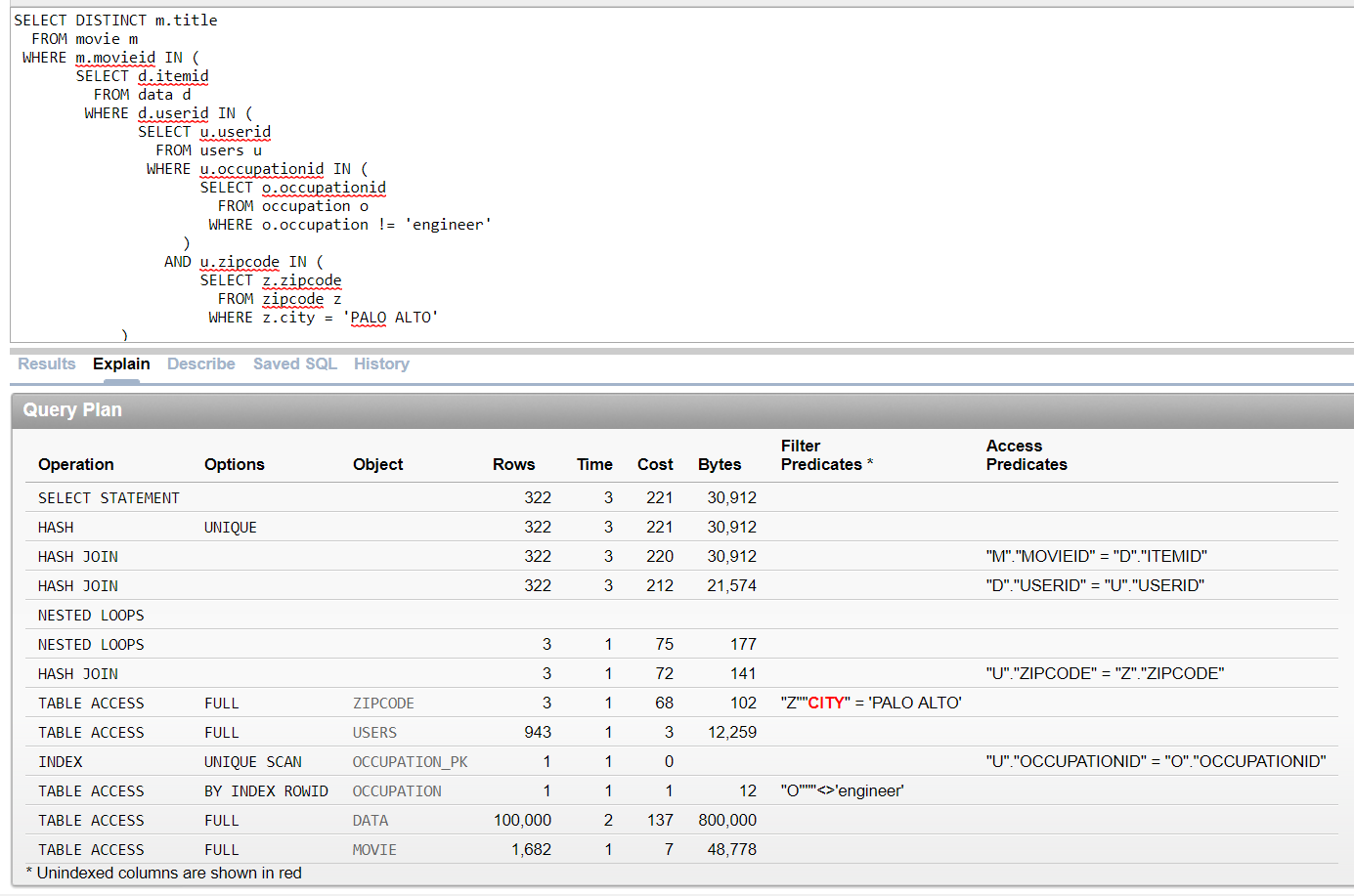
)

)

)







**SELECT city, property\_type, AVG(price) as avg\_price**

**FROM listings\_bigdata02**

**WHERE city='Queens'**

**GROUP BY property\_type, city;**

**SELECT property\_type, name, price**

**FROM listings\_bigdata02, calendar\_bigdata02**

**WHERE ID = listing\_ID**

**AND available = ‘t’ AND price > 50 AND price < 200**

**AND accommodates > 4 AND property\_type <> ‘House’;**

**SELECT \* FROM**

**(SELECT DISTINCT L.name, L.property\_type, L.price, N.neighbourhood\_group, L.number\_of\_reviews**

**FROM listings\_bigdata02 L, Neighbourhood N, Calendar C**

**WHERE L.ID = C.Listing\_ID AND N.neighbourhood = L.neighbourhood**

**AND C.available = ‘t’ AND N.neighbourhood\_group IN (‘Queens’,’Manhattan’)**

**AND L.price < 500 ORDER BY number\_of\_reviews DESC) X**

**WHERE rownum<=10;**

**SELECT host\_name, city, zipcode, property\_type, price**

**FROM listings\_bigdata02**

**WHERE City = ‘Bronx’**

**AND property\_type NOT IN (‘Apartment’,’House’)**

**AND room\_type LIKE ‘%room%’;**

**SELECT host\_name, city, zipcode, property\_type, price**

**FROM listings\_bigdata02**

**WHERE ID NOT IN ( SELECT L.ID**

**FROM listings\_bigdata02 L, neighbourhood N**

**WHERE L.neighbourhood = N.neighbourhood**

**AND N.neighbourhood\_group = ‘Bronx’ );**

**SELECT L.name, C.listing\_date,L.price**

**FROM listings\_bigdata02 L, reviews\_bigdata02 R, calendar\_bigdata02 C**

**WHERE L.ID=R.listing\_id AND L.ID=C.listing\_id**

**AND C.available='t‘ AND l.review\_scores\_accuracy>6**

**AND l.review\_scores\_rating>8**

**AND EXTRACT(YEAR FROM r.review\_date)=2017**

**GROUP BY l.name,c.listing\_date,l.price;**

**SELECT \* FROM listings\_bigdata02 L ,**

**(SELECT COUNT(R.listing\_id) AS count\_of\_listings, R.listing\_id**

**FROM reviews\_bigdata02 R**

**WHERE R.review\_date BETWEEN TO\_DATE('01-01-2010') AND TO\_DATE('09-01-2017') GROUP BY R.listing\_id) R2**

**WHERE L.ID=R2.listing\_id**

**AND l.review\_scores\_accuracy=9 AND count\_of\_listings>30;**

**SELECT ROW\_NUMBER() OVER ( ORDER BY review\_scores\_accuracy, number\_of\_reviews) AS rank, ID AS listing\_ID, name**

**FROM listings\_bigdata02**

**SELECT L.ID, L.name, count(R.listing\_ID), L.review\_scores\_accuracy, L.price**

**FROM listings\_bigdata02 L, reviews\_bigdata02 R**

**WHERE L.ID = R.listing\_ID AND L.ID IN (**

**SELECT LL.ID**

**FROM listings\_bigdata02 LL**

**WHERE LL.review\_scores\_accuracy > 8 )**

**GROUP BY L.name, L.ID, L.review\_scores\_accuracy, L.price**

**HAVING count(R.listing\_ID) >= 85**

**ORDER BY count(R.listing\_ID) DESC;**

**SELECT L.host\_name**

**FROM listings\_bigdata02 L**

**WHERE L.city = ‘Manhattan’**

**UNION ALL**

**SELECT L.host\_name**

**FROM listings\_bigdata02 L**

**WHERE L.city = ‘Brooklyn’;**

SQLS :

**SELECT :**

**Predicates :**

**Aggregates Natural JOIN :**

**Theta JOIN :**

**Multiple types of subqueries :**