Aggregation Questions

Attributes

|  |  |
| --- | --- |
| Natural language | Entity / Attributes |
| field | field\_production.field |
| production of {oil | petroleum},  {oil | petroleum} production | field\_production.oil |
| production of gas, gas production | field\_production.gas |
| year, yearly | field\_production.year |
| month | field\_production.month |
| state | field\_production.state |
| basin | field.basin |
| operator, company, enterprise | field.operator |
| contract | field.contract\_number |

Functions / Clauses

|  |  |
| --- | --- |
| Natural language | SQL |
| minimum, min | min |
| maximum, max | max |
| mean, average, avg | avg |
| count, number of, how many | count |
| sum, how much |  |
| by, per | GROUP BY followed by ORDER BY to be more human friendly |

1. What was the **average** **monthly** production of oil in state of Rio de Janeiro?

SELECT AVG(oil) as answer

FROM field\_production

WHERE state = 'Rio de Janeiro'

ANSWER

50,4695152881537

1. What was the **average monthly** production of oil in Rio de Janeiro?

It is the same question as above, but the state is implied. The only way to know that Rio de Janeiro is a state attribute data is to browse the data from all columns and see which one has the most occurrences of Rio de Janeiro. This is a KwS problem, but it is good to understand how it is solved for relational databases, in RDF databases seems to be a simpler problem since there are no differences in data and metadata.

SELECT AVG(oil) as answer

FROM field\_production

WHERE state = 'Rio de Janeiro'

ANSWER

50,4695152881537

1. What was the **average yearly** production of oil in state of Maranhão?

If the attribute is stored in the entity on a monthly basis and the question in natural language is asked on an annual basis, it is necessary to group by year and use the sum aggregation function on the attribute in the select clause before applying another aggregation function, if required by the question.

SELECT AVG(SUM(oil)) as answer

FROM field\_production

WHERE state = 'Maranhão'

GROUP BY year

ANSWER

630,66565

1. What was the **average** **monthly** oil production of the operator Petrobras?

Similar to question (1), the difference is that instead of using the state attribute, the operator attribute is used and it belongs to another table making necessary the use of a join, but this is a KwS problem.

SELECT AVG(oil)

FROM field\_production fp, field f

WHERE fp.field = f.name AND f.operator = 'Petrobras'

ANSWER

131,207750330333

1. What was the **mean** **monthly** gas production **per field**?

The special words “per” or “by” trigger the use of the GROUP BY clause followed by ORDER BY clause to be more human friendly. The attribute just after the “per” or “by” should be present in the SELECT, GROUP BY and ORDER BY clauses.

SELECT field, AVG(gas) AS gas\_production

FROM field\_production

GROUP BY field

ORDER BY field

FIELD GAS\_PRODUCTION

ABALONE 14,0848461538461

ACAJÁ-BURIZINHO 0,0106940476190476

ACAUÃ 0,00219285714285714

ÁGUA GRANDE 158,395617857143

... ...

1. What was the mean gas production **per month** **per field**?
2. What was the **per month** mean gas production **per field**?

The questions (6) and (7) have the same result set. They are similar to question (5), but they have two attributes grouped.

SELECT month, field, AVG(gas) AS gas\_production

FROM field\_production

GROUP BY month, field

ORDER BY month, field

MONTH FIELD GAS\_PRODUCTION

January ABALONE 0

January ACAJÁ-BURIZINHO 0,0108142857142857

January ACAUÃ 0,00225714285714286

January ÁGUA GRANDE 151,216828571429

… … …

1. What was the **per field** mean petroleum/gas production **per month**?

The order of attributes preceded by "per" or "by" matters

SELECT field, month, AVG(gas) AS gas\_production

FROM field\_production

GROUP BY field, month

ORDER BY field, month

FIELD MONTH GAS\_PRODUCTION

ABALONE January 0

ABALONE February 0

ABALONE March 9,7153166666667

ABALONE April 18,7371333333333

… … …

1. What was the mean monthly petroleum production by field in state of Rio de Janeiro?

The same style as question (5) but using a WHERE clause, it is a KwS problem.

SELECT field, AVG(oil) AS oil\_production

FROM field\_production

WHERE state = 'Rio de Janeiro'

GROUP BY field

ORDER BY field

FIELD OIL\_PRODUCTION

ALBACORA 0

ALBACORA LESTE 0

ANEQUIM 340,128963095238

ATAPU 44,4902666666667

… …

1. What was the average monthly production of oil per field in state of Rio de Janeiro and year 2016?

Similar to the previous one, but with two conditions in the WHERE clause, note that the second condition can use the key word "and" or "in" while the first one only "in". It is a KwS problem.

SELECT field, AVG(oil)

FROM field\_production

WHERE state = 'Rio de Janeiro' AND year = '2015'

GROUP BY field

ORDER BY field

FIELD OIL\_PRODUCTION

ALBACORA 0

ALBACORA LESTE 0

ANEQUIM 461,038833333333

ATAPU 44,4902666666667

… …

1. What was the average monthly production of gas before 2016 by operator?

Similar to the previous ones, what changes is a KwS problem.

SELECT operator, AVG(oil)

FROM field\_production

WHERE periodo < 2016

GROUP BY operator

ORDER BY operator

1. What was Petrobras’s **average** **monthly** gas production by field?

Similar to the previous ones, what changes is a KwS problem.

SELECT field, AVG(gas)

FROM field\_production

WHERE operator = 'Petrobras'

GROUP BY field

ORDER BY field

1. In how many years did Petrobras produce more than 1.000 cubic meters of oil per field?
2. In how many years did Petrobras produce between 1.000 and 10.000 barrels of oil per field?
3. In how many states did Petrobras produce less than 1.000 tonnes of gas?
4. How many fields produced more than than 1.000 cubic meters of oil per month?
5. Which operators produced less than 1.000 cubic meters of oil per month?
6. How many months was production below 1.000 cubic meters of oil per month?
7. In how many years was production between 1.000 and 10.000 barrels of oil per month?
8. How many operators produce between 1.000 and 10.000 barrels of oil per year?
9. In how many fields is Petrobras’s monthly production more than the average for all fields?