Scalable Data Engineering, Exercise 2

Henry Haustein

Task 1

(a) False, for some reporting functions we need `OVER`

(b) True

(c) False, this does `CUBE`

(d) True

(e) True

(f) False, SQL integer division returns an integer

(g) False, `ORDER BY` in an `OVER` clause has no effect on the sorting of the result. And an `ORDER BY` outside an `OVER` only sorts the result.

(h) False, MDX operates on data cubes. This can sometimes work on relational databases but it's not always the case.

Task 2

(a) SQL:

```sql
SELECT avg(ORDERS.O_TOTALPRICE) FROM ORDERS
```

	avg(ORDERS.O_TOTALPRICE)
1	151219.537631631

(b) SQL (I'm using Sqlite for this task, it doesn't support `ROLLUP`):

```sql
SELECT REGION.R_NAME,
FROM REGION, ORDERS, CUSTOMER, NATION
WHERE
  ORDERS.O_CUSTKEY = CUSTOMER.C_CUSTKEY AND
  CUSTOMER.C_NATIONKEY = NATION.N_NATIONKEY AND
  NATION.N_REGIONKEY = REGION.R_REGIONKEY
GROUP BY REGION.R_NAME
UNION
SELECT NULL, sum(ORDERS.O_TOTALPRICE) FROM ORDERS
```

	R_NAME	sum(ORDERS.O_TOTALPRICE)
1	NULL	226829306447.447
2	AFRICA	45230223902.2213
3	AMERICA	45306943255.2101
4	ASIA	45613415042.5609
5	EUROPE	45793265459.7115
6	MIDDLE EAST	44885458787.7611