Applied Databases

Data Cleaning Exercise Sheet

Get *salespeople.csv* from Moodle and load it into a table in MySQL with the following attributes:

- `id INTEGER`
- `name VARCHAR(50)`
- `role VARCHAR(50)`
- `sales_area VARCHAR(50)`
- `q1 DOUBLE(8,2)`
- `q2 DOUBLE(8,2)`
- `q3 DOUBLE(8,2)`
- `q4 DOUBLE(8,2)`

Question 1

All names should be in the format *firstname surname*.

Names currently not in that format are in the format *surname, firstname* and should be changed.

Question 2

An indicator column entitled *isNorthernMgr* should be added to the table with the value *1* if the salesperson is a *Manager* from sales_area *North*.

Question 3

Add a column entitled *Q1_Range* that contains the following values:

- <10K If the value for the *q1* column is less than 10,000
- <15K If the value for the *q1* column is between 10,000 and 14,999.99
- <20K If the value for the *q1* column is between 15,000 and 19,999.99
- <30K If the value for the *q1* column is between 20,000 and 29,999.99
- >30K If the value for the *q1* column is greater than 29,999.99.

Question 4

Create a view called *totals_view* that contains the unique *role* and a column called *total_sales* that contains the total of values in columns *q1, q2, q3* and *q4* for each role type.

```
mysql> SELECT * from totals_view;
+--------------------+-------------+
| role               | total_sales |
+--------------------+-------------+
| Salesperson        |   467467.91 |
| Senior Salesperson |   222002.01 |
| Manager            |   170532.11 |
+--------------------+-------------+
3 rows in set (0.00 sec)
```